



**AFRL-RQ-WP-TP-2017-0034**

**DESIGN OF EXPERIMENTS FOR MODEL CALIBRATION  
OF MULTI-PHYSICS SYSTEMS WITH TARGETED  
EVENTS OF INTEREST (PREPRINT)**

**Benjamin P. Smarslok**

**Hypersonic Sciences Branch  
High Speed Systems Division**

**Diane Villanueva**

**Universal Technology Corporation**

**MARCH 2017**

**DISTRIBUTION STATEMENT A: Approved for public release.  
Distribution is unlimited.**

*See additional restrictions described on inside pages*

**AIR FORCE RESEARCH LABORATORY  
AEROSPACE SYSTEMS DIRECTORATE  
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7542  
AIR FORCE MATERIEL COMMAND  
UNITED STATES AIR FORCE**

## NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the USAF 88th Air Base Wing (88 ABW) Public Affairs Office (PAO) and is available to the general public, including foreign nationals.

Copies may be obtained from the Defense Technical Information Center (DTIC)  
(<http://www.dtic.mil>).

AFRL-RQ-WP-TP-2017-0034 HAS BEEN REVIEWED AND IS APPROVED FOR  
PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//Signature//

BENJAMIN P. SMARSLOK  
Program Manager  
Hypersonic Sciences Branch  
High Speed Systems Division

//Signature//

MICHAEL S. BROWN, Branch Chief  
Hypersonic Sciences Branch  
High Speed Systems Division

//Signature//

JAMES H. MILLER, Principal Advisor  
High Speed Systems Division  
Aerospace Systems Directorate

This report is published in the interest of scientific and technical information exchange and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

\*Disseminated copies will show “//Signature//” stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b></p>						
1. REPORT DATE (DD-MM-YY) March 2017		2. REPORT TYPE Technical Paper Preprint		3. DATES COVERED (From - To) 01 October 2014 – 15 March 2017		
4. TITLE AND SUBTITLE DESIGN OF EXPERIMENTS FOR MODEL CALIBRATION OF MULTI-PHYSICS SYSTEMS WITH TARGETED EVENTS OF INTEREST (PREPRINT)				5a. CONTRACT NUMBER In-house		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER 61102F		
6. AUTHOR(S) Benjamin P. Smarslok (AFRL/RQHF) Diane Villanueva (Universal Technology Corporation)				5d. PROJECT NUMBER 3002		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER Q184		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Hypersonic Sciences Branch (AFRL/RQHF) High Speed Systems Division Air Force Research Laboratory, Aerospace Systems Directorate Wright-Patterson Air Force Base, OH 45433-7542 Air Force Materiel Command, United States Air Force				8. PERFORMING ORGANIZATION REPORT NUMBER AFRL-RQ-WP-TP-2017-0034		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Aerospace Systems Directorate Wright-Patterson Air Force Base, OH 45433-7542 Air Force Materiel Command United States Air Force				10. SPONSORING/MONITORING AGENCY ACRONYM(S) AFRL/RQHF		
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-RQ-WP-TP-2017-0034		
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A: Approved for public release. Distribution is unlimited.						
13. SUPPLEMENTARY NOTES PA Case Number: 88ABW-2016-3797; Clearance Date: 29 July 2016. The U.S. Government is joint author of this work and has the right to use, modify, reproduce, release, perform, display, or disclose the work.						
14. ABSTRACT <p>The design of hypersonic air vehicles involves coupled, multi-physics interactions, which are predicted through computational models of various levels of fidelity and accuracy. To reduce uncertainty and improve predictive capability, these models are calibrated with experimental data. Since the number of experiments is often limited, especially those conducted for structures undergoing the combined loading of hypersonic flight, optimal data collection is of great importance for uncertainty reduction and model validation. In this research, the maximum expected information gain is used to determine which wind tunnel specimen geometry, instrumentation locations, and observables are projected to be most informative for Bayesian calibration of the uncertain parameters of an aerothermal model. Higher fidelity simulations and synthetic experimental data are used to measure and compare the actual information gain from optimal designs to the expected information gain. It was observed that geometries and instrumentation locations at the limits of the design space provided the maximum expected information gain. Additionally, tests to measure the output of the furthest downstream model in the Bayesian network were favored due their ability to calibrate the full set of uncertain parameters. This study was extended to include an assumed cost model and a framework was built to trade-off cost and expected information gain.</p> <p>For accurate prediction of events of interest, the Targeted Information Gain for Error Reduction (TIGER) method is introduced to balance the placement of exploration points in the design space based on model accuracy and capturing the event of interest. This approach was compared to using sequential and all-at-once random data collection methods. The comparison of global and local prediction errors indicated that this is a feasible approach based on an analytical two-dimensional example. The method was also successful in a classification problem for flutter and critical limit cycle oscillation amplitude for a panel in hypersonic flow.</p>						
15. SUBJECT TERMS uncertainty quantification, validation, Bayesian techniques						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT: SAR	18. NUMBER OF PAGES 44	19a. NAME OF RESPONSIBLE PERSON (Monitor) Benjamin P. Smarslok	
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include Area Code) N/A	

# Design of Experiments for Model Calibration of Multi-Physics Systems with Targeted Events of Interest

Diane Villanueva<sup>1</sup> and Benjamin P. Smarslok<sup>2</sup>

*Air Force Research Laboratory, Wright-Patterson AFB, OH 45433*

The design of hypersonic air vehicles involves coupled, multi-physics interactions, which are predicted through computational models of various levels of fidelity and accuracy. To reduce uncertainty and improve predictive capability, these models are calibrated with experimental data. Since the number of experiments is often limited, especially those conducted for structures undergoing the combined loading of hypersonic flight, optimal data collection is of great importance for uncertainty reduction and model validation. In this research, the maximum expected information gain is used to determine which wind tunnel specimen geometry, instrumentation locations, and observables are projected to be most informative for Bayesian calibration of the uncertain parameters of an aerothermal model. Higher fidelity simulations and synthetic experimental data are used to measure and compare the actual information gain from optimal designs to the expected information gain. It was observed that geometries and instrumentation locations at the limits of the design space provided the maximum expected information gain. Additionally, tests to measure the output of the furthest downstream model in the Bayesian network were favored due to their ability to calibrate the full set of uncertain parameters. This study was extended to include an assumed cost model and a framework was built to trade-off cost and expected information gain.

For accurate prediction of events of interest, the Targeted Information Gain for Error Reduction (TIGER) method is introduced to balance the placement of exploration points in the design space based on model accuracy and capturing the event of interest. This approach was compared to using sequential and all-at-once random data collection methods. The comparison of global and local prediction errors indicated that this is a feasible approach based on an analytical two-dimensional example. The method was also successful in a classification problem for flutter and critical limit cycle oscillation amplitude for a panel in hypersonic flow.

## Introduction

The extreme environment of hypersonic flight leads an aircraft structure to exhibit highly coupled aerothermoelastic response. In order to effectively meet structural design margins and maximize aircraft performance by safely reducing design weight, uncertainty-quantified computational aero-thermal-structural models are necessary. This requires an understanding of the complex fluid-thermal-structural interactions of hypersonic flow and the uncertainties that hinder accurate modeling of aircraft structural response. Some of the sources of these uncertainties include imperfect knowledge of aerothermoelastic coupling, reduced-order model approximations, modeling assumptions, and limited data from experiments. With test data limited by experimental costs and the inability to fully replicate hypersonic environments through ground tests, the optimal design of model calibration experiments is of great importance.

Due to the cost and physical limitations of experimental studies, especially those conducted for structures undergoing the combined loading of hypersonic flight, the number and type of tests at even small scales (e.g., panel level) is limited. Examples of such tests include the aerothermal tests conducted by Glass and Hunt<sup>1</sup> in NASA's 8-foot High-Temperature Tunnel (HTT) on spherical domes protruding from a flat ramp subjected Mach 6.5 flow, which were designed to simulate a deformed hypersonic aircraft panel. In this paper, we seek to design aerothermal experiments similar to those performed by Glass and Hunt, focused on optimizing the geometry and instrumentation of a wind tunnel specimen for maximum uncertainty reduction in aerothermal models through calibration. Multiple observables can be measured in a wind tunnel test of this nature, namely aerodynamic pressure and heat flux. Due to experimental costs, it is assumed that there is a limited, discrete set of specimen and instrumentation locations available to be studied in a high-speed tunnel, such as the 8-foot HTT, under the desired hypersonic testing conditions.

In the first part of this study, a method is developed to obtain optimum data collection for calibration of coupled fluid-thermal-structural models for aerodynamic pressure and heating of rigid specimens with deformations corresponding to combinations of the first and second structural mode shapes subjected to hypersonic flow. The

---

<sup>1</sup> Research Engineer, Universal Technology Corporation

<sup>2</sup> Research Aerospace Engineer, Aerospace Systems Directorate, Structural Sciences Center

approach employed is to explore a combination of lower-fidelity models (i.e., 3<sup>rd</sup>-order piston theory and Eckert's reference temperature method) and corresponding discrepancy models to overcome computationally intractable coupled aerothermoelastic predictions from computational fluid dynamics (CFD). Inevitably, this combination of lower fidelity models and their discrepancy models are to some degree uncertain, such that a robust uncertainty quantification framework is necessary to ensure predictions at some confidence level. As part of an uncertainty quantification framework, calibration of these models with experimental data improves the accuracy and predictive capabilities.

As the number of experiments is limited, it is imperative to be able to anticipate the benefit from conducting experiments and maximize the amount of information that is gained from an experiment. Information theory and decision theory approaches have been used extensively in experimental design, taking different approaches in measuring the information gained from an experiment in terms of reducing the uncertainty in model parameters. For nonlinear relationships between observables and models, which is typical of aerothermoelastic models for hypersonics, these approaches include the maximization of expected information gain<sup>2-4</sup>, entropy<sup>5</sup>, and mutual information.<sup>6</sup> Bryant and Terejanu<sup>6</sup> sought to find the optimal sequence of experimental designs by maximizing the mutual information. Bayesian statistics provides a framework for integrating experimental observations and computational model predictions with uncertainty. Bayesian networks enable the fusion of various forms of information and capturing complex relationships between uncertainties and model predictions through nodes (i.e., conditional probabilities) of a network.<sup>7</sup> By incorporating experimental data into individual nodes, uncertainty can be reduced over the entire network. Bayesian networks are used in this study to represent interactions between aerothermoelastic models and experimental data for Bayesian model calibration. Previous work on Bayesian calibration of aerothermal models used historic data from the aforementioned Glass and Hunt wind tunnel experiments to quantify model discrepancy and input uncertainties.<sup>8-10</sup> This data set has been used for several purposes in recent and on-going research efforts, including aerothermal model calibration<sup>8,9</sup> and model validation studies.<sup>10,11</sup>

This work considers the maximization of the expected information gain criterion<sup>2,3,6</sup> to determine which design is optimal in terms of the geometry of the specimen and the instrumentation of the specimen. The expected information gain can be simply thought of as the expected change in prior to posterior distributions of uncertain parameters (measured by the Kullback-Leibler divergence) after an experiment is performed. Therefore, a large expected information gain value would correspond to a large change in the distribution of the uncertain parameters, which indicates that the experimental design provided information that had a large effect on the uncertainty. Expected information gain can be used to compare the utility of data gathered from experiments or simulations of various fidelities. Additionally, it allows the design of experimental conditions against cost (e.g., cost of experiment versus expected information gain).

For the aerothermal calibration experiment considered in the present study, the observables are aerodynamic pressure and heat flux at different locations on the specimen, with each specimen instrumented for either pressure or heat flux measurements. This allows for not only optimal design of the experimental specimen, but also the optimal instrumentation and measurement locations for the observables. The data collected can also be optimized such that only subsets of data in a multi-level or hierarchical system are measured.

With an experimental budget in mind, the second part of this study formulates the optimal data collection aerothermal problem to trade off the expected information gained of the experiment with the cost of the experiment itself. Using cost assumptions for instrumentation and manufacturing of the experimental specimen, the optimal designs are examined. The research provides the foundation of a framework to aid decision making for resource allocation for data collection for model calibration in a multi-physics system.

In addition to global accuracy, if a model is used to predict a specific event of interest, it is also important that some of the calibration data is collected near that region so the model can reliably and accurately predict the occurrence of the event. Some examples of such events include failure (e.g., exceeding a limit state corresponding to max stress or temperature, buckling) or transition into another regime (e.g., elastic to plastic, laminar to turbulent). Various approaches have been used to build surrogate models that improve probability of failure and limit state predictions<sup>20-24</sup> or locate feasible regions of the design space.<sup>25</sup> In communication theory and pattern recognition, information measures have been modified to include utilities based on the ability to meet a goal.<sup>26,27</sup> The value of the experiment was measured quantitatively by the amount of information added from the experiment, and qualitatively by its ability to meet a goal. In a similar manner, the third part of this study examines a weighted form of the EIG criterion where the weight is determined by the probability of the event of interest (EoI). However, to ensure global accuracy, data points are still selected globally based on a bi-objective formulation, called the Targeted Information Gain for Error Reduction (TIGER) criterion. The methodology is illustrated on a 2-D analytical example and for the identification of the Mach numbers for flutter and critical amplitude of limit cycle oscillation for a panel in hypersonic flow conditions.

As a whole, this research is part of an effort to incorporate data from experiments that consider subsets of physics (e.g., aerothermal response) in an uncertainty quantification framework for a multi-physics model. The major contributions of this work are

- An optimal data collection framework to allocate experimental resources to different types of experiments and models in a multi-physics system based on expected information gain
- Introducing budget constraints into an optimal data collection framework
- Developing a method to target events of interest into optimal data collection framework

The remainder of the report consists of three parts corresponding the contributions listed above.

## **Part 1: Optimal Design of Wind Tunnel Experiments for Aerothermal Model Calibration**

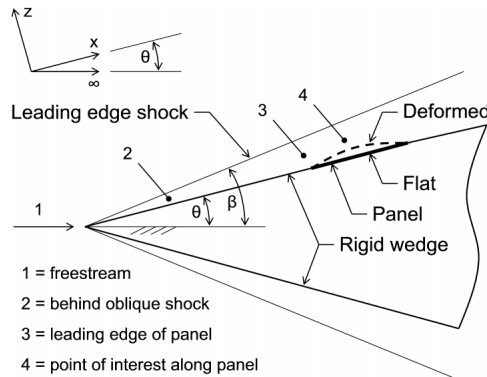
In the study described in Part 1, the maximum expected information gain is used to determine which wind tunnel specimen geometry, instrumentation locations, and observables are projected to be most informative for Bayesian calibration of the uncertain parameters of an aerothermal model. Higher fidelity simulations and synthetic experimental data are used to measure and compare the actual information gain from optimal designs to the expected information gain. A process of sequential data collection and model calibration is used until the maximum number of tests are reached. Of interest is the allocation of tests between models in the multi-physics systems, the information gained for the calibration parameters, and the reduction of uncertainty in the predictions of the quantities of interest.

Section I describes the aerothermal models used in this study and the experiments considered for model calibration. Section II details the optimal data collection methodology, and Section III illustrates these methods in the design of an experiment for aerothermal model calibration. The final section of Part 1 summarizes conclusions taken from this study.

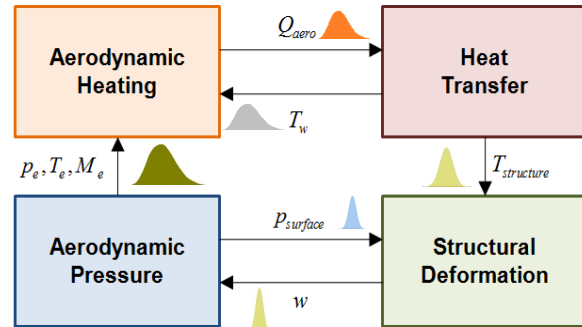
## **I. Aerothermal Model Definition and Experiments**

### **A. Aerothermoelastic Coupling**

Consider a panel section on the forebody of a representative hypersonic vehicle configuration, as shown in Figure 1.<sup>29</sup> As the vehicle is subjected to hypersonic flow (location '1'), an attached oblique shock is created at the forebody of the panel, which feeds back to alter the aerodynamic pressure on the panel. This is commonly referred to as the aeroelastic portion of the coupling. The panel is also subjected to aerothermal effects from aerodynamic heating. Naturally, this aerothermal component is coupled to the aeroelastic component, since a change in the temperature of the structure causes additional deformation, which in turn further alters both the aerodynamic pressure and the aerodynamic heating. Figure 2 schematically illustrates these fluid-thermal-structural interactions as a coupled aerothermoelastic response, including model components: aerodynamic pressure, aerodynamic heating, heat transfer, and structural deformation.<sup>30</sup>



**Figure 1. Representative panel behind shock on a hypersonic vehicle**



**Figure 2. Aerothermoelastic Coupling**

While Figure 2 represents the fully-coupled aerothermoelastic system, the present study will focus on the relationship between aerodynamic pressure and heating, known as the aerothermal portion of the model. In the

following subsection, a set of previous aerothermal experiments are described that are later used for calibration of the aerothermal model.

### B. Aerothermal Experiments

Glass and Hunt conducted hypersonic wind tunnel tests to investigate the thermal and structural loads on body panels in extreme environments.<sup>1</sup> This set of tests was designed to investigate the aerodynamic pressure and heat flux on a deformed panel in hypersonic flow. To simulate the deformed panel, rigid, spherical dome protuberances were constructed at different height-to-diameter ( $H/D$ ) ratios.

The protuberances are assumed to be rigid, such that the coupling between structural deformation and aerodynamic heating can be neglected. The coupled aerothermoelastic model in Figure 2 can be simplified to the aerothermal portion by examining only the aerodynamic pressure and aerodynamic heating.

Along with the Mach number ( $M_1$ ) and freestream pressure ( $p_1$ ) for each test run, the data reports both the aerodynamic pressure ( $p_4$ ) and aerodynamic heat flux ( $Q_4$ ) at the center of the flat plate and at 58 instrumented location on the spherical dome. For the purposes of this analysis in this paper, where only the panel behavior in two dimensions is considered, we limit the analysis to 11 points instrumented along the specimen's centerline.

The next section details the models that are used to estimate the pressure and heating.

### C. Aerothermal Models

Given the freestream flight conditions ( $p_1, M_1, T_1$ ) and the surface inclination angle ( $\theta$ ), the local conditions at the leading edge of the panel ( $p_3, M_3, T_3$ ) resulting from an oblique shockwave can be computed using oblique shock relations, shown in Eqs. (1) - (4). The oblique shock calculations do not have any dependency on the geometry of the panel itself, solely the surface inclination of the forebody and freestream conditions. Thus, it is only valid at locations where no structural deformation is present (i.e., a flat plate).

$$\frac{p_3}{p_1} = 1 + \frac{2\gamma}{\gamma+1}(M_1^2 \sin^2 \beta - 1) \quad (1)$$

$$\frac{\rho_3}{\rho_1} = \frac{(\gamma+1)M_1^2 \sin^2(\beta)}{(\gamma-1)M_1^2 \sin^2(\beta) + 2} \quad (2)$$

$$\frac{T_3}{T_1} = \frac{p_3 / p_1}{\rho_3 / \rho_1} \quad (3)$$

$$M_3^2 \sin^2(\beta - \theta) = \frac{M_1^2 \sin^2(\beta) + \frac{2}{\gamma-1}}{(\frac{2}{\gamma-1})M_1^2 \sin^2(\beta) - 1} \quad (4)$$

Once the flow properties at the leading edge of the panel are calculated from oblique shock relations, piston theory provides a simplified relationship between the unsteady pressure on the panel and turbulent surface pressure.<sup>12</sup> This simple pressure model is desired for computational tractability and uses the leading edge conditions to approximate the aerodynamic pressure load chord-wise across the panel ( $p_4, M_4, T_4$ ). In piston theory, the pressure prediction is dependent on the slope of the panel ( $\partial w / \partial x$ ) and the velocity of deformation ( $\partial w / \partial t$ ). As stated previously, the panel is assumed to be rigid hence  $\partial w / \partial t$  is zero. In the case of no deformation (i.e., a flat plate where  $\partial w / \partial x = 0$ ), the pressure across the panel is the same as the pressure at the leading edge, i.e.,  $p_3^{fp} = p_4^{fp}$ . A 3rd-order expansion of piston theory is presented in Eq. (5).

$$p_4 = p_3 + 2 \frac{q_3}{M_3} \left[ \left( \frac{1}{U_3} \frac{\partial w}{\partial t} + \frac{\partial w}{\partial x} \right) + \frac{\gamma+1}{4} M_3 \left( \frac{1}{U_3} \frac{\partial w}{\partial t} + \frac{\partial w}{\partial x} \right)^2 + \frac{\gamma+1}{12} M_3 \left( \frac{1}{U_3} \frac{\partial w}{\partial t} + \frac{\partial w}{\partial x} \right)^3 \right] \quad (5)$$

After calculating the aerodynamic pressure and flow conditions along the panel surface, the aerodynamic heat flux is predicted using the computationally efficient Eckert's reference temperature method assuming a calorically perfect gas.<sup>13</sup> The Eckert's reference temperature is computed by Eq. (6) and the heat flux across the spherical dome follows in Eq. (7).

$$T^* = T_3 + 0.5(T_w - T_e) + 0.22(T_{aw} - T_3) \quad (6)$$

$$Q_4 = St^* \rho^* U_e c_p^* (T_{aw} - T_w) \quad (7)$$

Where,  $St^*$  is the reference Stanton number,  $\rho^*$  is the reference density,  $U_e$  is the inviscid flow velocity at the dome location,  $c_p^*$  is the reference specific heat,  $T_{aw}$  and  $T_w$  are the adiabatic wall and actual wall temperature, respectively, and  $T_e$  is the boundary layer edge temperature at any location along the dome.

## II. Optimal Data Collection for Bayesian Model Calibration

In this section, a framework to optimally collect data for calibration of the aerothermal model is presented. First the Bayesian network, which includes model inputs, model discrepancy, measurement errors, and calibration parameters, is described. Then we introduce the metric to optimally collect data, the expected information gain of the future experiment.

### A. Bayesian Model Calibration

The Bayesian network in Figure 3 represents the aerothermal models and relationships described in Section I, where the observations  $y$  are data from a Glass and Hunt type experiment. Specifically, it shows the relationships between the aerodynamic pressure and heat flux model predictions ( $p_4$ ,  $Q_4$ ), Glass and Hunt data ( $y_{p4}$ ,  $y_{Q4}$ ), deterministic model inputs ( $p_1$ ,  $M_1$ ,  $T_1$ ,  $T_w$ ), measurement errors ( $\varepsilon_{yp}$ ,  $\varepsilon_{yQ}$ ), and discrepancy terms for calibration ( $\varepsilon_{PT}$ ,  $\varepsilon_{ERT}$ ). The model discrepancy terms will be discussed in the following subsection.

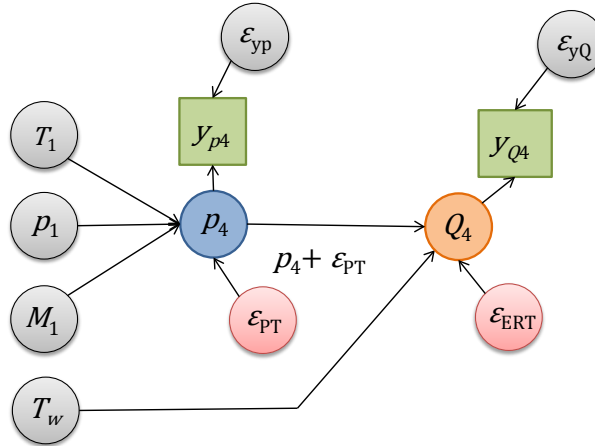


Figure 3. Aerothermal Bayesian Network

Related efforts<sup>8,9</sup> focusing on quantifying model uncertainty in aerothermal predictions followed the framework described by Kennedy and O'Hagan<sup>14</sup> that relates the measurement from the experiment  $y$  to the model output  $G$  through model discrepancy  $\varepsilon_G$  (also referred to as model error or model inadequacy), and measurement error  $\varepsilon_y$ . This relationship is shown as

$$y = G(x, \theta) + \varepsilon_G(x) + \varepsilon_y(x) \quad (8)$$

where  $x$  represents the design variables,  $\theta$  represents the uncertain parameters. In this work, the only uncertain parameters are contained in the discrepancy model (i.e., no uncertain parameters in the piston theory or Eckert's reference temperature models) such that this relationship becomes



$$y = G(x) + \varepsilon_G(x, \theta) + \varepsilon_y(x) \quad (9)$$

For the Glass and Hunt setup described previously, the discrepancy models  $\varepsilon_G$  are  $\varepsilon_{PT}$  and  $\varepsilon_{ERT}$ . The uncertain parameters  $\theta$  are defined in Sec. II.

Bayesian model calibration is used to obtain the distributions of the uncertain model parameters  $\theta$ , given observations of the pressure  $y_p$  and  $y_Q$ . Bayes' theorem is written for the aerothermal models and Glass and Hunt data in Eq. (10), where the prior distributions of the uncertain parameters are given by  $\pi(\theta)$ .

$$\pi(\theta | y_p, y_Q) = \frac{\Pr[y_p, y_Q | \theta] \pi(\theta)}{\int \Pr[y_p, y_Q | \theta] \pi(\theta) d\theta} \quad (10)$$

Bayesian model calibration incorporates all of the available data from the corresponding observables. In the next section, a metric will be introduced that will help quantify the significance of incorporating data into the Bayesian network, which will enable determination of optimal data collection.

### B. Expected Information Gain

For experiments aimed at parameter inference or Bayesian model calibration, a popular metric for measuring the utility of an experiment is the information gain. An optimal experimental design can be found through maximization of the expected information gain. Lindley<sup>1</sup> proposed the measure of the expected information gain as the expected Kullback-Leibler divergence<sup>15</sup> of the posterior (i.e., post-experiment) and prior distribution of the uncertain parameters. The expected information gain can be expressed as

$$U(x) = \int \int_{\Theta} \Pr(\theta | y, x) \ln \left[ \frac{\Pr(\theta | y, x)}{\Pr(\theta)} \right] d\theta \Pr(y | x) dy \quad (11)$$

where  $\theta$  denotes the uncertain parameters,  $y$  are the future experimental result, and  $x$  are the design variables.<sup>3</sup> This formulation can be interpreted intuitively by examining the possible effect of a future result,  $y$ . For example, let  $y$  decrease the entropy of  $\theta$  by a large amount (e.g., increasing  $\Pr(\theta | y, x)$  relative to the prior distribution of  $\Pr(\theta)$ ). In this case, the information gain is large and is thus more informative for inference or calibration. The expectation is taken over the possible experimental results  $y$  given a design  $x$ , prior distributions  $\Pr(\theta)$ , and measurement uncertainty  $\varepsilon_y$  of the experiment.

Monte Carlo sampling can be used to estimate the expected information gain in Eq. (11), as shown in Eq. (12).<sup>3,16</sup>

$$U(x) \approx \frac{1}{N} \sum_{i=1}^N \left\{ \ln[\Pr(y^{(i)} | \theta^{(i)}, x)] - \ln \left( \frac{1}{M} \sum_{j=1}^M \Pr(y^{(i)} | \theta^{(j)}, x) \right) \right\} \quad (12)$$

The Monte Carlo approximation for expected information gain that considers measurements at  $k$  locations on the specimen is shown in Eq. (13), where  $y_l$  represents the measurement at the  $l^{\text{th}}$  location of the set of candidate locations  $L$ .

$$U_{test}^{MCS}(x) = \frac{1}{N} \sum_{i=1}^N \left\{ \ln \left[ \prod_{l=1}^k (\Pr(y_l^{(i)} | \theta^{(i)}, x)) \right] - \ln \left( \frac{1}{M} \sum_{j=1}^M \left( \prod_{l=1}^k \Pr(y_l^{(i)} | \theta^{(j)}, x) \right) \right) \right\} \quad (13)$$

Similarly, the expected information gain for a single instrumentation location,  $l$ , is

$$U_{inst,l}^{MCS}(x) = \frac{1}{N} \sum_{i=1}^N \left\{ \ln [\Pr(y_l^{(i)} | \theta^{(i)}, x)] - \ln \left( \frac{1}{M} \sum_{j=1}^M \Pr(y_l^{(i)} | \theta^{(j)}, x) \right) \right\} \quad (14)$$

### C. Sequential Data Collection

The process of calibration of the uncertain parameters proceeds in iterations. In each iteration, the optimal geometry is found, followed by the optimal instrumentation locations. This is a sequential “greedy” approach that purely looks to exploit areas of large expected information gain in each iteration. In this study, the optimal geometry  $d$  belongs to a discrete set of candidate designs  $D$ . Additionally, the instrumentation locations  $d_{inst}$  belong to the discrete set of candidate locations  $L$ . The locations that are instrumented have a  $U_{inst}$  value larger than the second smallest  $U_{inst}$  value, which effectively leaves out the two locations with the smallest  $U_{inst}$  values. This can be summarized in the optimization problem formulation in Eq. (15).

$$\begin{aligned}
 \max_{d, d_{inst}} \quad & U_{max} = \max(U_{test}^P(d), U_{test}^Q(d)) \\
 \text{s.t.} \quad & d \in D \\
 & d_{inst} \subset L \\
 & U_{inst}(d_{inst}^i) > x_{(2)}(U_{inst}(L))
 \end{aligned} \tag{15}$$

Once the optimum geometry and instrumentation locations are found, a test is performed and the models are calibrated as described in Sec. II.B. This study uses Kriging surrogates of Reynolds Averaged Navier-Stokes computational fluid dynamics predictions developed by Crowell et al.<sup>18</sup> as synthetic test data. A small amount of noise is placed on the surrogate predictions. This noise was modeled as a normal distribution with a zero mean and standard deviation of 5% of the measurement standard deviation, which is discussed in the following section.

Figure 4 displays a flowchart of the process used to find the optimal geometry and instrumentation. In this study, the stopping criterion is the number of tests that are performed. Once the maximum number of tests is reached, the assumption is made that future tests will be performed for model validation rather than calibration. The area of the design of validation tests will be a subject of future research.

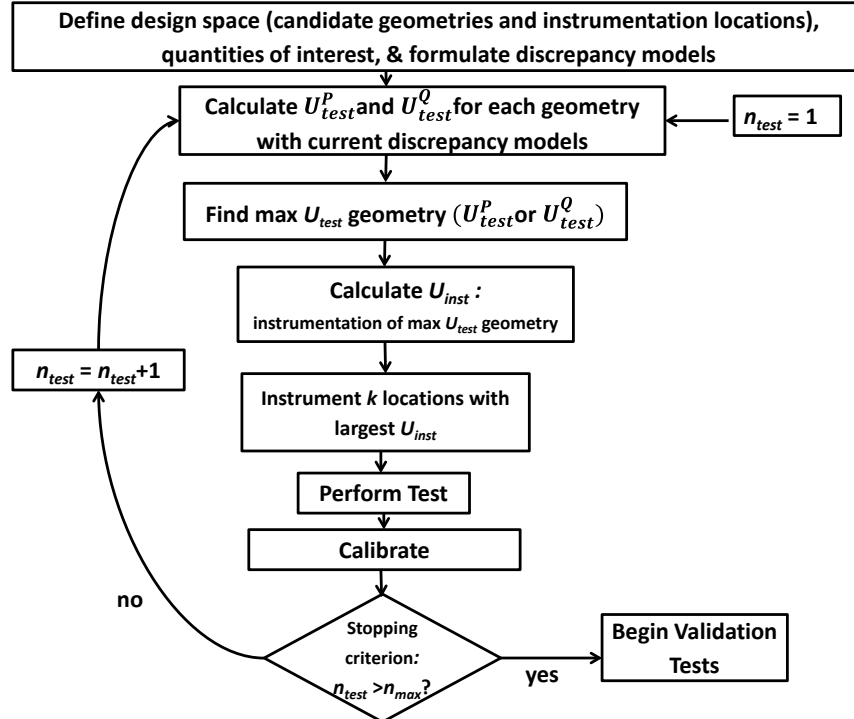


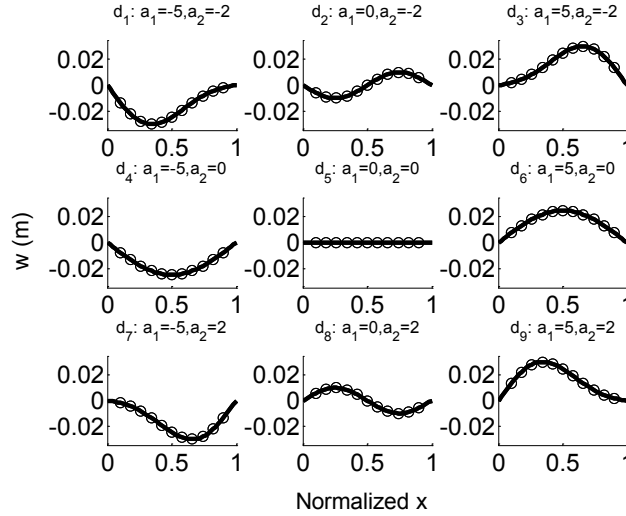
Figure 4. Flowchart of the process of finding the optimal geometry and instrumentation for calibration

### III. Optimal Data Collection for Aerothermal Experiments

The previous sections described a framework to select the optimal geometry, instrumentation, and type of test for sequential Bayesian calibration of aerothermal models. Next, the expected information gain methodology is applied to discrepancy models for aerodynamic pressure and heat flux with candidate geometries and instrumentation locations. Figure 5 displays all nine candidate geometries with the displacement  $w$  over the normalized panel length. Eleven candidate locations for instrumentation are also shown. The displacement is a function of the scale factors  $a_1$  and  $a_2$  for the first and second mode shapes  $\Phi$ , respectively, where  $x$  is the coordinate along the length of the geometry.

$$w(x) = a_1\Phi_1(x) + a_2\Phi_2(x) \quad (16)$$

The bounds  $a_1$  are  $[-5, 5]$ , and  $[-2, 2]$  for  $a_2$  give displacements that are close to the magnitudes of the heights of the spherical domes used by Glass and Hunt and used in previous model calibration studies.<sup>8,9</sup> Eleven candidate instrumentation locations were considered, which corresponds to the number that Glass and Hunt considered. The 11 candidate locations were equally spaced between 10% and 90% of the specimen length. Figure 5 also displays these candidate instrumentation locations on each geometry. In this study, 9 of the 11 candidate locations will be instrumented based on ranking of  $U_{inst}$  values for the optimal geometry.



**Figure 5. Candiate geometries, where candidate locations for instrumentation are represented by unfilled circles**

Table 1 lists the test conditions: freestream Mach number, temperature, pressure, and forebody surface angle of inclination (as shown in Figure 1). Additionally, the table lists the measurement uncertainties.

**Table 1. Test conditions**

Condition	Value
Mach number, $M_1$	8
Temperature, $T_1$	226.5 K
Pressure, $p_1$	1197 N/m <sup>2</sup>
Surface inclination angle	5 degrees
$\varepsilon_{yp}$	$N(0, 500)$ N/m <sup>2</sup>
$\varepsilon_{yQ}$	$N(0, 8000)$ W/m <sup>2</sup>

The discrepancy model for piston theory  $\varepsilon_{PT}$  is a linear function of the scale factors of the first two mode shapes ( $a_1$  and  $a_2$  for the first and second mode shape, respectively) and slope of the dome protuberance  $\frac{\partial w}{\partial x}$ , while Eckert's reference temperature  $\varepsilon_{ERT}$  is modeled as a linear function of the slope, as shown in Eqs. (17) and (18).

$$\varepsilon_{PT} = b_0 + b_1 a_1 + b_2 a_2 + b_3 \frac{\partial w}{\partial x} \quad (17)$$

$$\varepsilon_{ERT} = c_0 + c_1 \frac{\partial w}{\partial x} \quad (18)$$

Table 2 lists the prior distributions for the uncertain parameters (i.e., the coefficients of discrepancy models). All prior distributions are modeled as uniform distributions.

**Table 2. Prior distributions for uncertain model discrepancy parameters**

Parameters	Prior Distribution
$b_1$	$U(-500, 500)$
$b_2$	$U(-15e3, 15e3)$
$b_3$	$U(-15e3, 15e3)$
$b_4$	$U(-10e3, 10e3)$
$c_0$	$U(1e4, 5e4)$
$c_1$	$U(1e5, 1e6)$

As in Eqs. (13) and (14), Monte Carlo sampling was used to estimate the expected information gain. Here, 1,000 samples for  $N$  and  $M$  were used in this estimate. Slice sampling,<sup>19</sup> a form of Markov Chain Monte Carlo, was used for calibration to 2,000 samples from the posterior distributions with 20 burn-in samples.

The total number of tests to be performed was set at six. In the first iteration of the optimal data collection and calibration process, the maximum expected information gain geometries and instrumentation for heat flux and pressure were both found; both tests were performed and used for the first calibration. The remaining iterations only select one geometry for either pressure or heat flux for testing and calibration. Therefore, a total of 5 iterations provides the desired number of 6 test results. Twenty realizations of this optimal data collection and calibration process were performed to assess the effect of randomness. Thus, there are 20 distinct sets of tests, optimal geometries, instrumentation locations, and calibrated predictions of the pressure and heat flux.

The remainder of this section will examine one realization (Realization #3) to show trends in the selection of the test type, geometry, and instrumentation locations, along with their effect on the prediction of pressure and heat flux. Finally, a general trends from all twenty realizations are summarized.

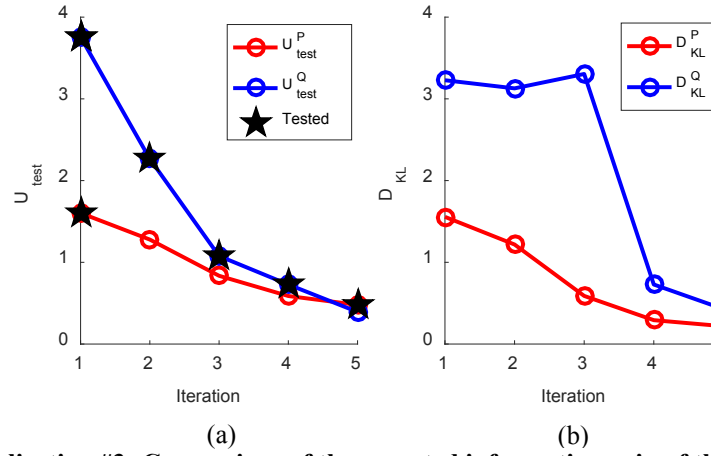
#### A. Results for a Single Realization

A description of the type of test and geometry in each experimental design iteration is shown in Table 3. Heat flux tests outnumbered pressure tests, which only occurred at the first iteration (as required) and last iteration.

**Table 3. Description of tests for Realization #3**

Iteration	Design	Type	$a$	$a_2$
1	$d_1$	$Q$	-5	-2
	$d_7$	$p$	-5	2
2	$d_9$	$Q$	5	2
3	$d_7$	$Q$	-5	2
4	$d_3$	$Q$	5	-2
5	$d_3$	$p$	5	-2

Figure 6 displays the maximum  $U_{test}$  in each iteration for each type of test. The expected information gain for the heat flux test is larger than a pressure test, leading to heat flux tests until the  $U_{test}$  from a pressure test is larger in the final iteration.

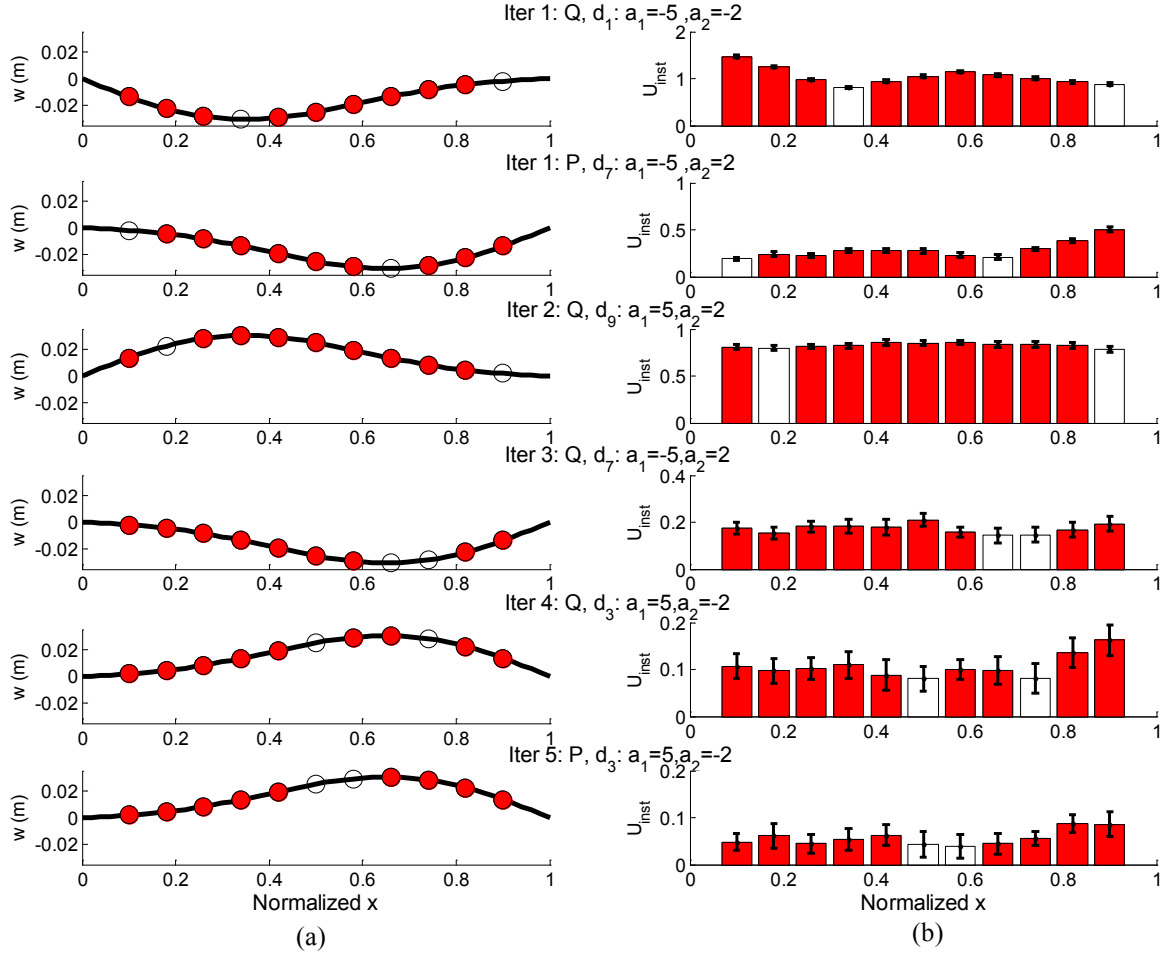


**Figure 6. Realization #3: Comparison of the expected information gain of the test, and the KL divergence of prior and posterior.**

The Kullback-Leibler divergence  $D_{KL}$  from prior to posterior distributions of the uncertain parameters was also calculated at each iteration as a measure of the actual information gain. For this calculation of  $D_{KL}$ , the uncertain parameters were treated together as a multivariate normal distribution. Comparing Figs. 7a and 7b, the tests led to a change in the distribution of the uncertain parameters following the trend of what was expected from the estimate of  $U_{test}$ .

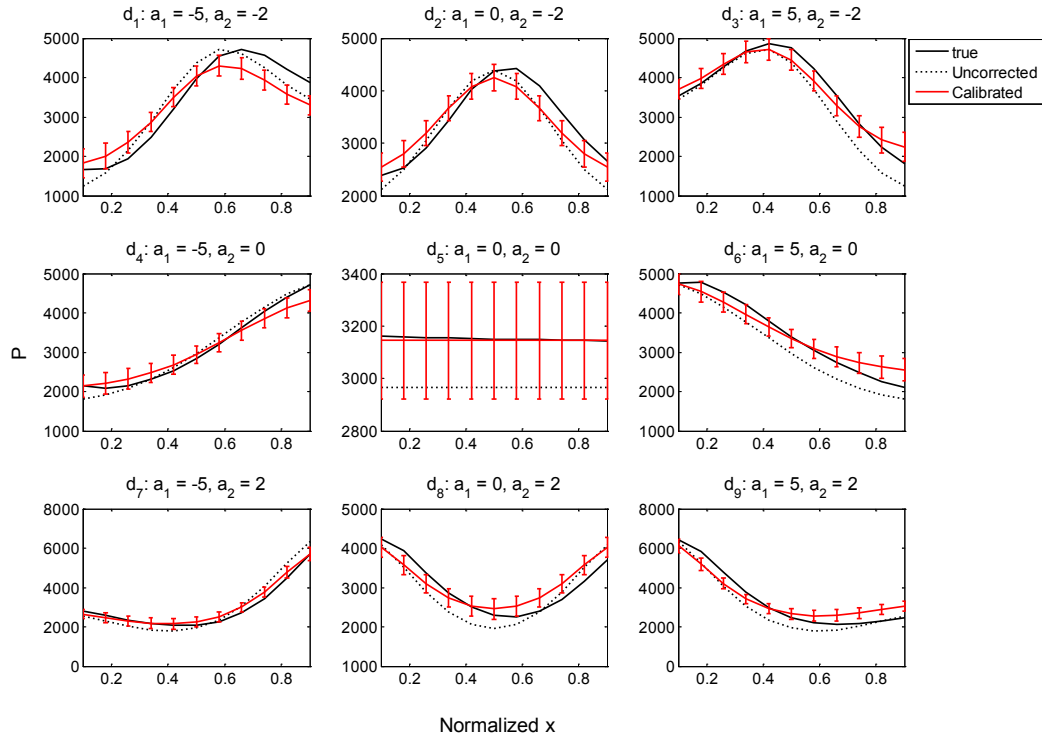
The larger  $U_{test}$  values for  $Q$  tests compared to  $p$  tests in Figure 6a can be attributed to the pressure prediction feeding into the heat flux prediction as shown in the Bayesian network from Figure 3. This setup results in  $Q$  measurements that calibrate both  $\varepsilon_{PT}$  and  $\varepsilon_{ERT}$ , whereas  $p$  measurements only calibrate  $\varepsilon_{PT}$ . Therefore,  $U_{test}^Q$  is larger until the uncertainty in the parameters of  $\varepsilon_{ERT}$  is much smaller compared the parameters of  $\varepsilon_{PT}$ .

The optimal instrumentation of the design iterations is shown in Figure 7a. The expected information gain value of each instrumentation location  $U_{inst}$  is shown in Figure 7b. Also shown are error bars at  $\pm\sigma$  around each  $U_{inst}$  value, which were found by conducting 10 separate Monte Carlo estimates of  $U_{inst}$ . A general trend shown by the optimal instrumentation locations is withholding instrumentation at locations where the slope of the specimen geometry is near zero. For example, the specimens for the  $Q$  and  $p$  tests in the first iteration are not instrumented at the point where the displacement is greatest (i.e., where the slope is near zero) and the end of the geometry where the displacement levels off to zero. However, noise in the calculation from Monte Carlo sampling does cause some instrumentation to go against this trend. For example, in Iteration 4, the points around the maximum displacement/minimum slope points are not instrumented. At this point in the iterations, the  $U_{inst}$  are all relatively small compared to the first few iterations, such that the error from the Monte Carlo estimate is large compared to  $U_{inst}$ .

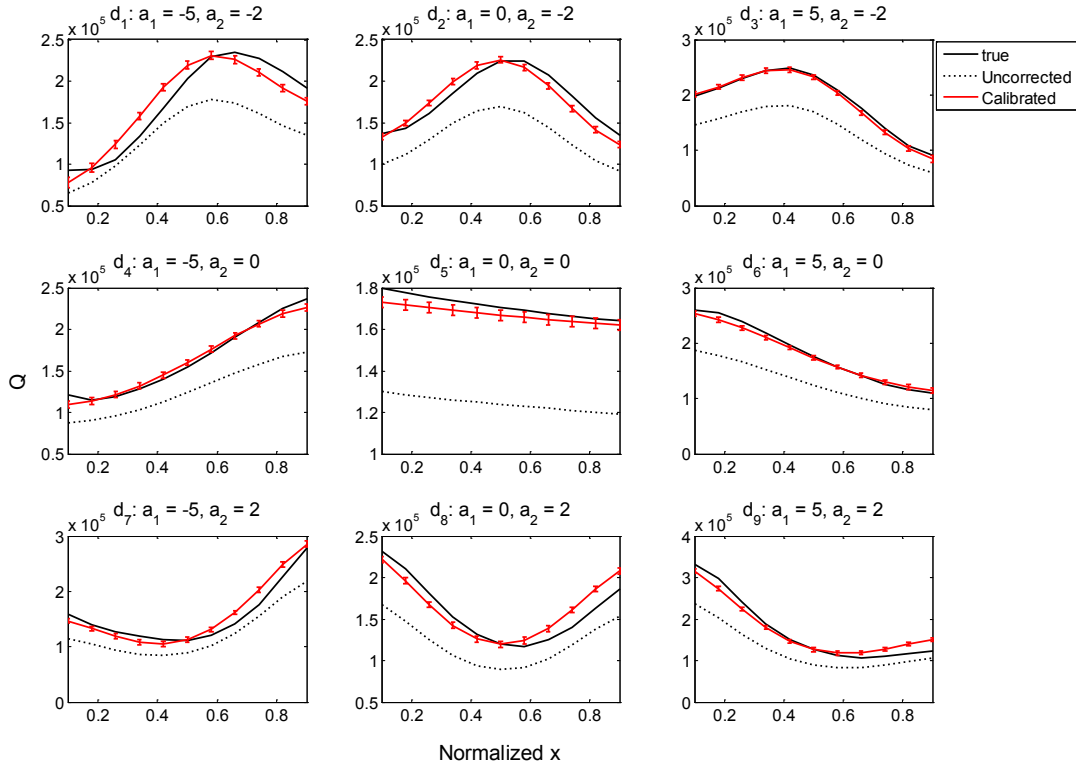


**Figure 7. Realization #3: (a) Instrumentation of the optimal geometry in each iteration with instrumented points are shown as filled red circles, and (b)  $U_{inst}$  values with error bars corresponding to  $\pm\sigma$ , which result from 10 repetitions of each  $U_{inst}$  calculation**

Figure 8 and Figure 9 compare the calibrated predictions of pressure and heat flux, respectively, of all candidate designs after five iterations to the uncorrected predictions and the “true” value from the CFD surrogates. For the pressure predictions, the reduction in uncertainty was most notable for the two geometries that underwent both pressure and heat flux tests, designs  $d_3$  ( $a_1 = 5, a_2 = -2$ ) and  $d_7$  ( $a_1 = -5, a_2 = 2$ ). This improvement is most evident in the areas where the slope is large, which is at the rear of the specimen for both designs. The improvement in the heat flux predictions is much more apparent, as shown in Figure 9. In most cases, the uncorrected prediction required a simple shift (i.e., bias) of the prediction, which was mostly captured in the calibration of the coefficient  $c_0$ .



**Figure 8. Realization #3: Comparison of nominal true, uncorrected, and calibrated pressure predictions. The calibrated prediction shows the  $\pm 2\sigma$  bounds on the prediction.**



**Figure 9. Realization #3: Comparison of nominal true, uncorrected, and calibrated heat flux predictions. The calibrated prediction shows the  $\pm 2\sigma$  bounds on the prediction.**

### B. Overall Trends for 20 Realizations

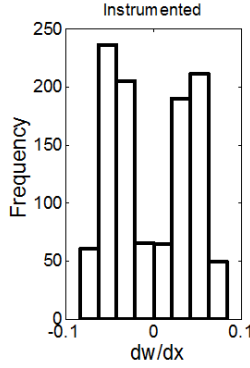
Considering all 20 realizations with 6 tests each, the percentage of tests performed on the candidate designs is displayed in Table 4. The tests were distributed among the geometries that were combinations of the first and second mode shape ( $d_1$ ,  $d_3$ ,  $d_7$ , and  $d_9$ ). This occurs because  $\varepsilon_{PT}$  is a function of both  $a_1$  and  $a_2$ , so data for a geometry that is purely mode 1 ( $d_3$  and  $d_6$ ), purely mode 2 ( $d_2$  and  $d_8$ ), or flat ( $d_5$ ), would not calibrate  $b_1$ ,  $b_2$ , or both  $b_1$  and  $b_2$ , depending on the geometry.

**Table 4. Geometries chosen for test for 20 realizations**

Geometry	Percentage of Tests (%)
$d_1$	27.5
$d_3$	22.5
$d_7$	24.2
$d_9$	25.8
$d_2, d_4, d_5, d_6, d_8$	0

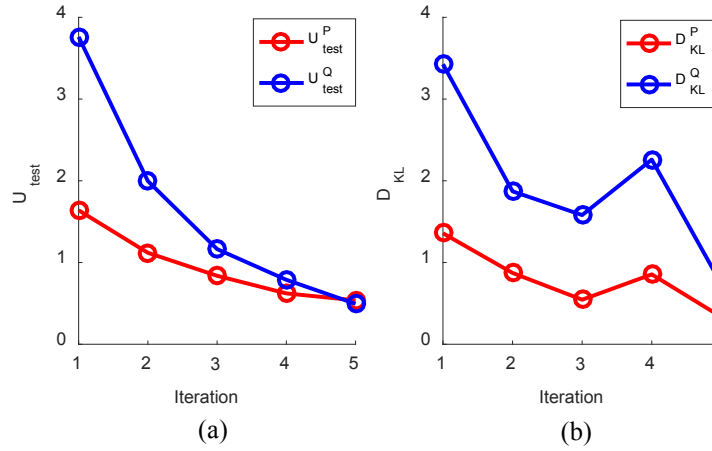
A total of 1,320 locations were examined when considering all 20 realizations of 6 tests and 11 candidate instrumentation locations each. As shown in Figure 10, most instrumented locations had slope values away from zero. Since both  $\varepsilon_{PT}$  and  $\varepsilon_{ERT}$  are functions of the slope, locations with a slope near zero provide less information than the points with higher slope.





**Figure 10. Histogram of slopes of instrumented locations over 20 realizations of 6 tests and 11 candidate instrumentation locations for each test**

Realizations 3 had larger expected information gain values from heat flux tests, and this was observed when looking at the mean  $U_{test}$  over the 20 realizations. This is displayed in Figure 11, along with the mean KL divergence. Note that there is a large increase in mean  $D_{KL}$  at iteration 4, which is indicative of test data that shifted the mean, variance, or both of the uncertain parameter distributions. Furthermore, since this is a mean value, it is affected by large values of  $D_{KL}$  that occur in later iterations.



**Figure 11. Mean  $U_{test}$  and  $D_{KL}$  of 20 realizations**

As another point of comparison, the error of each prediction after 5 iterations was compared against the “true” value obtained from the CFD surrogates. Additionally, the errors compared to those obtained when all data was used in at all-at-once calibration. When all data is used, this corresponds to 198 data points (both pressure and heat flux tests, 9 candidate geometries, and 11 instrumentation locations each). In comparison the sequential EIG method only used 54 data points, or approximately 27% of the data points. The comparison over the median root mean square error  $e_{RMS}$  normalized by the range of the true values over 20 realizations of each method is shown in Table 5. It was observed that the sequential EIG method provided nearly equal accuracy as using all possible data.

**Table 5. Comparison of median normalized  $e_{RMS}$  over 20 realizations**

	Median $e_{RMS}$	
	All data	Sequential EIG
$p$	0.070	0.076
$Q$	0.064	0.063

Over the 20 realizations, twice as many heat flux tests occurred as pressure tests. The iterations where aero-pressure tests were potentially chosen were at the first iteration, as dictated by the prescribed test plan, and the final iteration. This occurs because the  $Q$  measurement is further downstream in the Bayesian network of Figure 3, and can thus be used to calibrate the uncertain parameters of both  $\varepsilon_{PT}$  and  $\varepsilon_{ERT}$ . Therefore, in general, more information can be gained from heat flux tests.

#### IV. Summary

A framework based on expected information gain was developed to determine the optimal design of experiments to maximize the reduction in model uncertainty through Bayesian model calibration. This investigation focused on the design of the geometry and instrumentation of hypersonic wind tunnel specimens for calibration of aerothermal models. The conditions for the experiment were based off of the historic tests performed by Glass and Hunt in NASA's 8-ft high temperature wind tunnel; however, the geometry of the specimens was designed as a function of structural mode shapes. Due to the costs of hypersonic wind tunnel tests, it was assumed that the number of tests and instrumentation locations of the specimen was limited. Therefore, an optimization problem was formulated to determine which type of test, geometry, and instrumentation locations would be most informative for model calibration. A sequential, greedy approach was taken to iteratively locate the design with the maximum expected information gain and use synthetic test data to calibrate the uncertain parameters until the maximum number of allowable tests was reached. The specimen configurations that combined mode shapes and instrumentation locations with large slopes provided the largest expected information gain values due to the form of the discrepancy model. Additionally, it was observed that heat flux tests occurred more frequently than pressure tests because measurements of heat flux data affect the calibration uncertain parameters in both piston theory and Eckert's reference temperature method. This occurs because the heat flux model is the furthest downstream model in the aerothermal Bayesian network.

Overall, this research is part of an effort in the development of an uncertainty quantification framework for hypersonic aircraft structures. The challenges include determining how to most effectively incorporate data from a subset of the coupled physics, and how to assess the confidence in predictions when it is required to extrapolate across multiple, individually-validated, coupled physics.

### **Part 2: Budgeting Model Calibration Experiments with Expected Information Gain**

In Part 2, we extend the previous study to include cost assumptions for the different types of tests (i.e., pressure and heat flux) and geometries of the specimen. Furthermore, the instrumentation types are combined on a single specimen. That is, based on expected information gain, each candidate location will be instrumented for either pressure or heat flux rather than single specimens of all pressure or all heat flux instrumentation. We also consider optimal design of a batch of experiments, comparing a single test two a batch of two or three tests.

The first section of Part 2 introduces a cost model. Section II formulates the optimization problem to trade-off expected information gain from the test, and examined the Pareto front for maximum information gain and minimum cost. Section III examines the optimal design of a batch of experiments, comparing the information gain from two more expensive tests to two cheaper tests for the same cost.

#### I. Cost Model

For the costs of the experiment, we first assume that more complex geometries are more difficult to manufacture. Therefore, we penalize more complex geometries by forming a cost model that is a function of the mode scale factors  $a_1$  and  $a_2$  where  $A = [a_1, a_2]^T$ . The cost of the geometry for a test is

$$C_A(A) = 3 \left( \frac{|a_1|}{5} + \frac{|a_2|}{2} \right) + 12 \quad (19)$$

Next, we make the assumption that heat flux tests are 3 times more expensive than a pressure test. The instrumentation type is represented as a binary vector  $S$  for which each element represents an instrumentation, where 0 represents a pressure measurement and 1 represents a heat flux measurement (again, note that this study combines pressure and heat flux instrumentation on a single specimen). Therefore, the instrumentation cost is

$$C_S(s) = \#\{S : s = 0\} + 3(\#\{S : s = 1\}) \quad (20)$$

The total cost of the experiment  $C_{total}$  is then the sum of the geometry and instrumentation costs.

$$C_{total}(X) = C_A(A) + C_S(S) \quad (21)$$

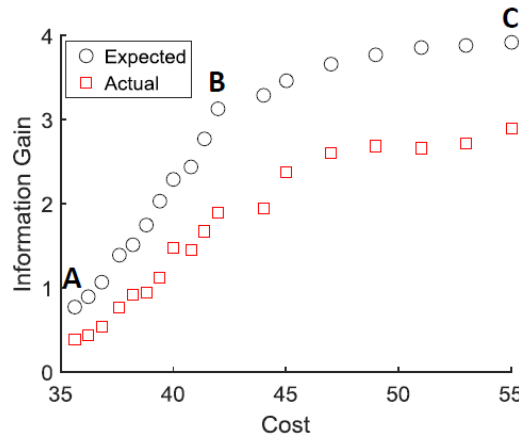
## II. Optimization to Trade Off Expected Information Gain and Cost

The cost models of the previous section are used as part of a bi-objective optimization problem to maximize the expected information gain and minimize the cost of the test. This formulation is represented as

$$\begin{aligned} & \max_X \text{EIG}(X), \min C_{total}(X) \\ & \text{where} \\ & X = [A^T, S^T]^T \\ & A = [a_1, a_2]^T \\ & S = [s_1, s_{11}]^T \\ & -5 \leq a_1 \leq 5 \\ & -2 \leq a_2 \leq 2 \\ & s := \begin{cases} 0 & \text{if pressure} \\ 1 & \text{if heat flux} \end{cases} \end{aligned} \quad (22)$$

The optimization problem was solved with the MATLAB function *gamultiobj*. The total number of design variables is 13, with two variables for the two mode scale factors and 11 for the instrumentation locations. Note that *gamultiobj* does not handle integer variables as required by this study, so the Pareto optimal designs given by *gamultiobj* are rounded to the nearest integer.

The resulting Pareto Front is shown in Figure 12. For each design on the Pareto Front, the actual information gain was calculated by performing a test (using synthetic data as described in Part 1) and calibrating by Bayesian calibration. The actual information gain is the KL divergence between the prior and posterior distributions of the uncertain parameters, approximating the posterior as a multivariate normal distribution.



**Figure 12. Pareto Front to maximum EIG and minimum cost. The expected and actual information gains are compared.**

The geometries and instrumentation types of Designs A, B, and C shown in Figure 12 were examined. Figure 13 shows that with increasing cost the mode shapes become more complex and more resources were allocated to instrument for heat flux. For example, Design A is a shallow mode 1 type deformation with a cost of 35.6 and EIG of 0.78. Increasing the cost to 42 as in Design B, results in a design that has a larger mode 1 scale factor and EIG of 3.13. Finally, the most expensive design, Design C, is a combination of the two mode shapes with the largest magnitudes of  $a_1$  and  $a_2$  and more heat flux instrumentation locations.

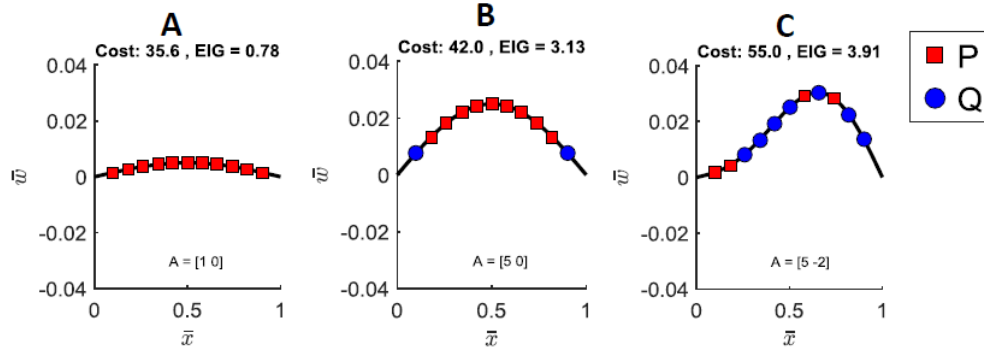


Figure 13. Geometry and instrumentation of three designs on the Pareto Front.

Next, the optimization problem was solved for a batch of two and three designs. This was achieved by adding the additional design variables to the design variable vector of Eq. (22). Therefore, there are 26 and 39 design variables for a batch of two and three designs, respectively. Solving the problem with *gamultiobj* results in the Pareto Fronts shown in Figure 14. The actual information gain for each Pareto optimal design is also displayed. The results from a single test are also shown as a point of comparison.

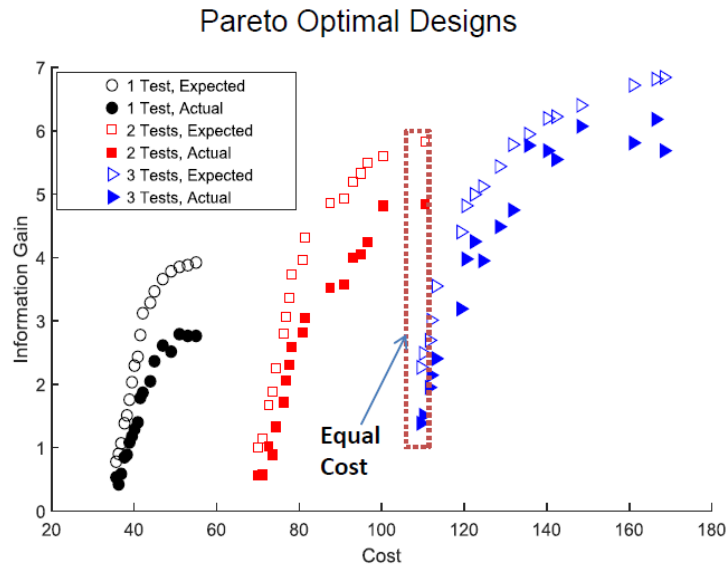


Figure 14. Pareto Front for maximum EIG and minimum cost for a batch sizes of 1, 2, and 3 tests.

The two optimal designs with a nearly equal cost as shown in Figure 14. The comparison was made between two more expensive tests and three cheaper tests as shown in Figure 15 and Figure 16, respectively. The two more expensive tests were combination of the two mode shapes and measured heat flux at most locations, whereas the three cheaper designs were either mode 1 deformations or a flat plate with mostly pressure instrumentation. It was observed that both the expected and actual information gain was larger for the batch of two more expensive designs at a nearly equal cost.

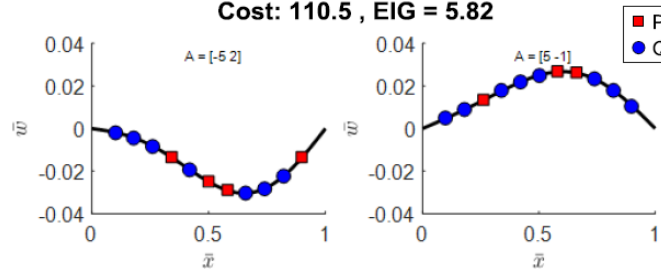


Figure 15. A Pareto optimal design for a batch of two tests used for the equal cost comparison

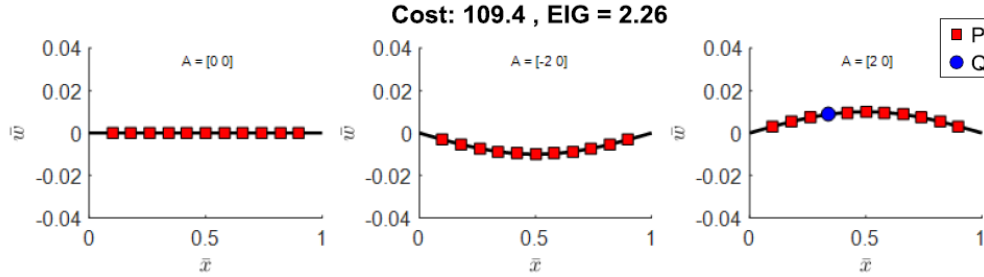


Figure 16. A Pareto optimal design for a batch of three tests used for the equal cost comparison

### III. Summary

This research lays the foundation for a framework to include cost assumptions in an optimization problem to trade off the information gain from a test and the cost of performing the test. For the aerothermal problem, it was observed that more complex geometries (i.e., larger combinations of mode scale factors) and heat flux instrumentation occurred with increasing cost and information gain. Studies such as these will help aid decision makers in designing tests for model calibration when the number of tests is limited due to cost.

### Part 3: Targeting Events of Interest in Experimental

While Parts 1 and 2 were concerned with obtaining data to calibrate models with the goal of global accuracy, to capture specific events, such as failure, optimal data collection methods can aid in achieving models that also predict targeted events. In Part 3, the Targeted Information Gain for Error Reduction (TIGER) method uses expected information gain to balance the placement of exploration points in the design space based on model accuracy and capturing the event of interest. This approach was compared to using sequential and all-at-once random data collection methods. The comparison of global and local prediction errors indicated that this is a feasible approach based on an analytical two-dimensional example. The method was also successful in a classification problem for flutter and critical limit cycle oscillation amplitude for a panel in hypersonic flow.

#### I. Introducing Targeted Events of Interest into Optimal Data Collection Framework

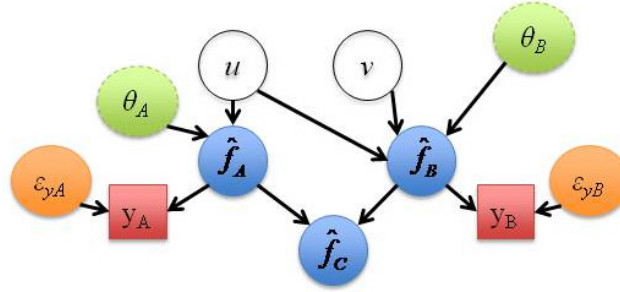
As in Parts 1 and 2, the Kennedy and O'Hagan framework was used to relate experimental data to the model output through model discrepancy and measurement uncertainty. The reader is referred to Part 1 for a more thorough description of this framework and the uncertainties. Additionally, Parts 1 and 2 both used the expected information gain (EIG) criterion to design calibration experiments. This section introduces a weighted variant of the expected information gain ( $EIG_w$ ), along with the TIGER formulation that balances the global accuracy benefit of EIG and the targeting ability of the  $EIG_w$ .

Gaussian process (GP) surrogate models are used in this study, which provide a flexible way to include data points with measurement uncertainty (i.e., noisy data). Rather than having a separate model  $G$  and discrepancy model  $\varepsilon_G$  (as shown in Eq. (8)), the uncertain parameters are contained in the GP. The GP takes the form

$$\hat{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \mathbf{b} + Z(\mathbf{x}) \quad (23)$$

where  $\hat{f}$  represents that the model is an approximate model. Both terms of the GP model contain uncertain parameters:  $\mathbf{h}(\mathbf{x})^T \mathbf{b}$  is the mean function,  $\mathbf{b}$  is a vector of coefficients, and  $Z$  is a stationary Gaussian process with zero mean and covariance function. The uncertain parameters  $\theta$  for the GP model are thus the coefficients of the mean function and parameters of the covariance function. Note that the number of uncertain parameters would vary for the choice of mean function and covariance function. For a more detailed description of the GP surrogate models the reader is referred to Rasmussen and Williams (2005).<sup>29</sup>

Consider a system of three models, where models  $\hat{f}_A$  and  $\hat{f}_B$  feed into  $\hat{f}_C$ , for which the probability of  $\hat{f}_C$  exceeding a limit state is the event of interest. Experimental data  $y$  can be obtained for  $\hat{f}_A$ ,  $\hat{f}_B$ , and  $\hat{f}_C$ , or data may be limited to a subset of the models. Additionally, each experimental measurement  $\varepsilon_y$  may have different levels of uncertainty. In the example Bayesian network in Figure 17,  $u$  and  $v$  are the design variables, where  $v$  is shared between models  $\hat{f}_A$  and  $\hat{f}_B$ , where  $\hat{f}_A$  is only a function of  $u$ . For this scenario, data may be obtained for all models except  $\hat{f}_C$ . This example will be re-visited in the illustrative example in Sec. II.



**Figure 17. Bayesian network representation of the relationship between design variables, models, uncertain parameters, experimental data, and measurement uncertainties**

#### A. Weighted Expected Information Gain

In order to target regions of the design space where an event of interest (EoI) occurs,  $EIG$  can be weighted by the probability of the EoI. This is achieved by introducing the probability into the Monte Carlo estimate of  $EIG$ , such that this weight is calculated for every realization of the set of  $\theta$ . Therefore, if the probability of the EoI is small when  $EIG$  is large, the weighted criterion would drive optimal designs away from that area of the design space. The weighted expected information gain estimate  $EIG_W$  is shown in Eq. (24).

$$EIG_W(x) \approx \Pr(EoI(x, \theta)) \frac{1}{N} \sum_{i=1}^N \left\{ \ln[\Pr(y^{(i)} | \theta^{(i)}, x)] - \ln \left( \frac{1}{M} \sum_{j=1}^M \Pr(y^{(i)} | \theta^{(j)}, x) \right) \right\} \quad (24)$$

Note that  $\Pr(EoI(x, \theta))$  does not need to be the actual probability of the EoI itself. For example, if it is set as the probability of failure, the designs chosen by  $EIG_W$  may be driven to areas where the estimated probability is 1. This may result in inaccurate estimation of the entire failure region, particularly if the failure region is large, because points are only placed where the probability approaches 1. Instead, this probability may be set as the probability of being near the limit state of the EoI. For example, for an EoI defined by  $g \geq \tau$ , the term  $\Pr(EoI(x, \theta))$  can be replaced by  $\Pr(g - 2\tau \leq \kappa \leq g + 2\tau)$  where  $\tau$  defines an area about the limit state.

#### B. Targeted Information Gain for Error Reduction

Using the  $EIG_W$  criterion alone as a measure of utility would drive the optimal design to areas of predicted high probability of the EoI. Therefore, we introduced a bi-objective formulation, the Targeted Information Gain for Error Reduction (TIGER) formulation, where the utility function consists of both  $EIG$  and  $EIG_W$ . The optimization problem of Eq. (25) is used to find the optimal design  $x^*$ , where  $\alpha$  weights the two objectives.

$$\begin{aligned} \max_x U_{TIGER} &= \alpha EIG(x) + (1 - \alpha) EIG_W(x) \\ \text{s.t. } x &\in X \end{aligned} \quad (25)$$

The purpose of  $\alpha$  is to balance global accuracy through designs chosen with large  $EIG$  and local, near EoI accuracy through those with large  $EIG_W$ . When only maximizing  $EIG_W$ , there is a risk of missing other areas of the design space where the EoI might occur. This is possible if one or more of the models is inaccurate in the design space. A natural way to avoid the use of test points to determine model error is to use a cross-validation metric, here the partial prediction error sum of squares  $PRESS_{RMS}$ . This is found by leaving out a design point, re-training with the remaining data, and measuring the error at that point to get  $e_{XV}$  at that point. The operation is repeated for  $p$  training points to form a vector of  $e_{XV}$ . The  $PRESS_{RMS}$  is calculated by

$$PRESS_{RMS} = \sqrt{\frac{1}{p} e_{XV}^T e_{XV}} \quad (26)$$

For each model, the  $PRESS_{RMS}$  is used to determine the value of  $\alpha$ . The value of  $\alpha$  is increasing with  $PRESS_{RMS}$ , such that a large  $PRESS_{RMS}$  corresponds to a large value of  $\alpha$ . The exact form of  $\alpha$  as a function of  $PRESS_{RMS}$  is user-defined, but the authors propose a linear function as used in the example problems in this paper.

The process of calibration of the uncertain parameters proceeds in iterations. In each iteration, the optimal design is found and added to the data set, and the models are calibrated with the new data. The flowchart in Figure 18 displays the sequential data collection process using the *TIGER* criterion for  $n$  models. When there are multiple models and corresponding experimental responses that can be collected, it is possible to compare the utility of collecting each form of data. Two branches are present in the flowchart to account for multiple experiments in an iteration (e.g., obtain  $y_A$  and  $y_B$  to calibrate  $\hat{f}_A$  and  $\hat{f}_B$ , respectively) or a single experiment in an iteration (e.g., obtain  $y_A$  or  $y_B$  to calibrate  $\hat{f}_A$  or  $\hat{f}_B$ , respectively). In the example shown in Figure 17, this would mean that the utilities of  $y_A$  and  $y_B$  would determine which experiment was performed to obtain calibration data as in the “single experiment per iteration” branch. For the same example, in an alternate scenario, both experiments could be performed at the different  $x^*$  to calibrate  $\hat{f}_A$  and  $\hat{f}_B$ , as in the “multiple experiments per iteration” branch. The methodology described is applicable to either scenario.

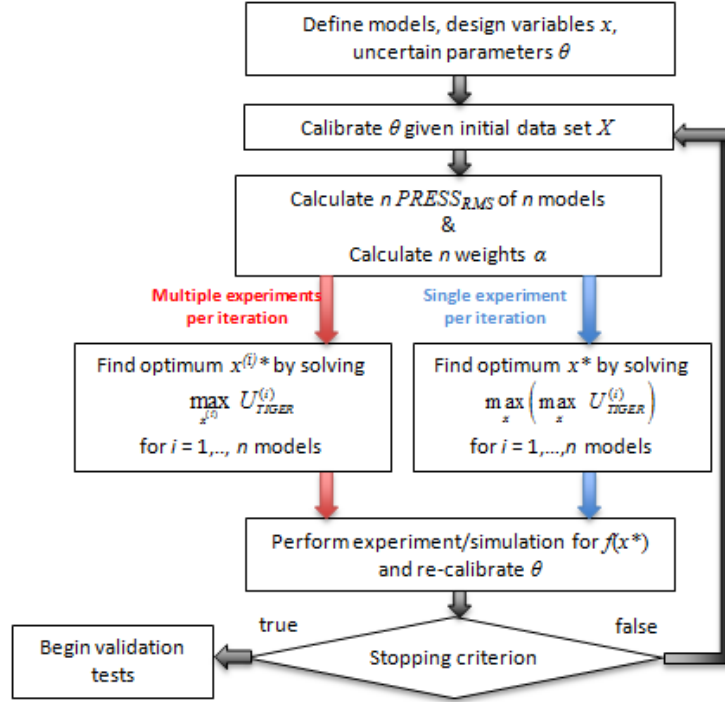


Figure 18. Flowchart of sequential data collection process using *TIGER* criterion.

A global optimization algorithm, such as DIRECT<sup>31</sup>, can be used to find the optimum. The DIRECT algorithm proceeds by iteratively dividing the design space into hyperrectangles, focusing on areas where the objective function is promising while also exploring the design space. In the following section, the DIRECT algorithm is used for a two-dimensional analytical example. In the application problem, the design space is divided into a discrete set at which the objective function is evaluated at all candidate locations from which the optimum is chosen.

## II. Illustrative Example

The sequential data collection methods described in the previous section are illustrated on a two-dimensional analytical function. The Targeted Information Gain for Error Reduction (*TIGER*) method is compared against using the objective functions of its two constituent parts: 1) sequential *EIG*, and 3) sequential *EIG<sub>W</sub>*. The results are compared to two all-at-once design of experiments with calibration, full factorial (FF) and Latin Hypercube Sampling (LHS).

### A. Problem Description

The modified camelback function from Picheny et al.<sup>22</sup> is a single function that was separated into three models for the purposes of this study. The function is described by Eqs. (27)–(31), where the design space is  $[-1, 1]$  for design variables  $u$  and  $v$ . The event of interest is the failure of the response  $f_C$ , where failure occurs if  $f_C \geq 1.3$ . Note that  $f_A$  is a one-dimensional function of  $u$ , while  $f_B$  is a function of both design variables. The Bayesian network shown in Figure 17 represents this example.

$$f_A(u) = \left( 4 - 2.1\bar{u}^2 + \frac{1}{3}\bar{u}^4 \right) \bar{u}^2 \quad (27)$$

$$f_B(u, v) = \bar{u}\bar{v} + (-4 + 4\bar{v}^2) \bar{v}^2 \quad (28)$$

$$f_C(u, v) = f_A(u) + f_B(u, v) - 0.7 \quad (29)$$



$$\bar{u} = 1.2u - 0.1 \quad (30)$$

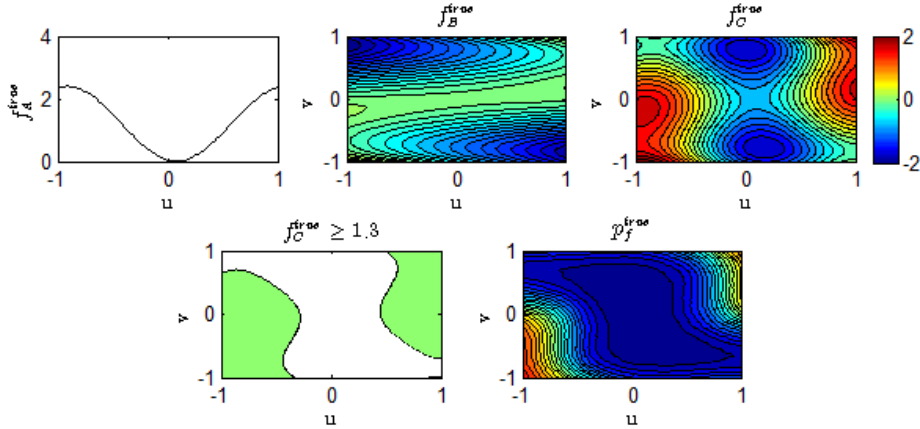
$$\bar{v} = 0.9v \quad (31)$$

**Table 6. Distributions of uncertain parameters for illustrative example**

Parameter	Distribution
$u$	$N(\mu_u, 0.28)$
$v$	$N(\mu_v, 0.28)$
$\varepsilon_{yA}$	$N(0, 0.01)$
$\varepsilon_{yB}$	$N(0, 0.02)$

For this problem, GP surrogates are built to provide predictions of  $\hat{f}_A$  and  $\hat{f}_B$ , and data collection experiments can only be performed to calibrate the  $\theta_A$  and  $\theta_B$  (epistemic uncertainties). The measurement uncertainties  $\varepsilon_{yA}$  and  $\varepsilon_{yB}$  are known. Additionally, the design variables have some aleatory uncertainty that is known. The distributions of all random (aleatory) uncertainties are provided in Table 6. The Bayesian network shown in Figure 17 represents this example.

Figure 19 displays the contours of the nominal true values of each model and the probability of failure  $p_f$ . Failure regions where  $f_C$  exceeds 1.3 are in the upper right and lower left corners of the design space.



**Figure 19. True values of nominal  $f_A, f_B, f_C$ , failure region ( $f_C \geq 1.3$ ), and  $p_f$**

In this work, the GP surrogates are calibrated with the available data, but the distributions of uncertain parameters  $\theta_A$  and  $\theta_B$  are not explicitly obtained. Rather, the GP surrogates are trained by maximizing the likelihood of  $\theta$ , such that a deterministic value is obtained. The GP surrogates in this study have a zero mean function and use the Matérn covariance function.<sup>29</sup> Random realizations of the predictions  $\hat{f}$  are obtained by taking the mean value of the prediction and sampling from the prediction uncertainty (i.e, the prediction standard deviation) to represent model uncertainty of the GP surrogate. Note that an alternative is sampling from the posterior distribution of  $\theta$  through Markov Chain Monte Carlo to obtain realizations of the predicted  $\hat{f}$ .

When the  $PRESS_{RMS}$  of the GP surrogates are large, then points are added in regions with large unweighted  $EIG$ . In this example, each model  $\hat{f}_A$  and  $\hat{f}_B$  has a corresponding  $\alpha$ , which is bounded between  $[0, 1]$  and is linear between  $PRESS_{RMS}$  values between 0 and 0.15, as shown in Figure 20. The exact relationship between  $PRESS_{RMS}$  and  $\alpha$  is subjective and can be tuned as necessary based on intuition, design space exploration, or level of ‘aggression’ in locating the event of interest.

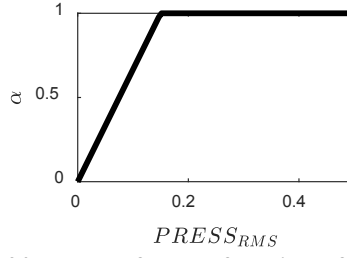


Figure 20. Value of  $\alpha$  as a function of  $PRESS_{RMS}$

### B. Comparison of Data Collection Methods

The sequential data collection with the Targeted Information Gain for Error Reduction (TIGER) methodology was compared to two cases: 1) sequential data collection with maximum  $EIG$  criterion 2) sequential data collection with maximum  $EIG_W$  criterion. Additionally, comparisons were made to all-at-once design of experiments with full factorial sampling and Latin Hypercube sampling. All data collected (i.e.,  $y_A$  and  $y_B$ ) is ‘synthetic’ data obtained by evaluating Eqs. (27)-(31), given a randomly sampled value of the uncertain parameters in Table 6. That is, a random sample of  $u$  and  $v$  are taken given the nominal  $\mu_u$  and  $\mu_v$  from the data collection method, and random samples of the measurement uncertainties  $\varepsilon_{yA}$  and  $\varepsilon_{yB}$  are added to the computed  $f_A$  and  $f_B$ .

All sequential data collection methods used an initial DoE of five points found by Latin Hypercube Sampling. Therefore, initially there was a combined total of 10 training points for the two approximations  $\hat{f}_A$  and  $\hat{f}_B$ .<sup>3</sup> Note that for  $\hat{f}_A$ , this puts two points at  $u = -1$  and  $u = 1$ . In each iteration, an optimum point was found by comparing the maximization criterion for  $\hat{f}_A$  and  $\hat{f}_B$ , and the models are calibrated with experimental data for the optimum design for either  $\hat{f}_A$  or  $\hat{f}_B$ . This corresponds to the blue “single experiment per iteration” branch in the flowchart in Figure 18. Forty iterations were performed where one point was added per iteration, leading to a total of 50 points for the final prediction.

The weight for the  $EIG_W$  criterion is determined by the probability of being near the limit state, represented by  $\tau$ , in the space of the response. The weight is determined by  $\Pr(1.3 - 2\tau \leq \hat{f}_C \leq 1.3 + 2\tau)$ , where  $\tau = 0.15$  in this study. Here, the random uncertainties in inputs  $u$  and  $v$  are not accounted for, such that the probability is determined only by the prediction and prediction uncertainty of the surrogate. It is cheap to calculate this probability because it is done using the surrogate approximations of  $\hat{f}_A$  and  $\hat{f}_B$  with 2,500 Monte Carlo samples obtained from the GP prediction and prediction uncertainty (equivalently, one could sample from uncertain distribution of  $\theta_A$  and  $\theta_B$ ). Note that accounting for the random uncertainty in  $u$  and  $v$  would require double loop sampling from the distribution of epistemic uncertainties  $\theta$  in the outer loop, and random uncertainties in the inner loop. Since  $f_A$  is only a function of  $u$ , the weight for  $EIG_W$  for performing an experiment for  $y_A$  is determined by the maximum probability at  $u$  taking  $v$  at 100 uniformly spaced points between  $[-1, 1]$ .

For this study, the stopping criterion of the DIRECT optimizer was a maximum number of function evaluations of 250. Calculating TIGER from Eq. (6) involved estimating of  $EIG$  and  $EIG_W$  with  $N = M = 1,000$  samples in Eq. (24). The probability of failure was obtained using 2,500 samples of  $u$  and  $v$  given the distributions of their random uncertainties in Table 6.

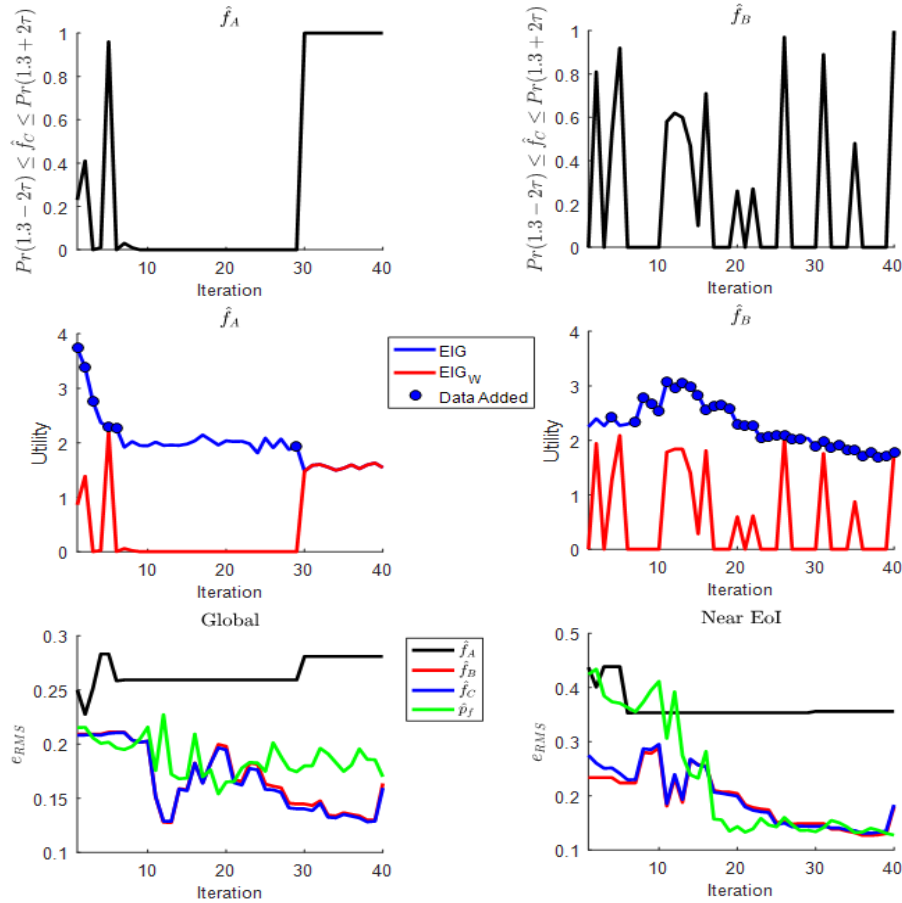
The following subsections compare the results obtained from 20 realizations (e.g., repetitions) for each data collection method to observe the general trends and to account for randomness in each method from Monte Carlo sampling and uncertain input parameters. In addition, a graphical comparison is provided of the points added by each data collection method for a single realization. This allows the reader to compare the placement of points in the design space obtained by the different data collection methods. Those results are accompanied by a comparison of the errors

<sup>3</sup> In this study, GP surrogates are built and require an initial set of training points. In an alternate scenario, a model, physical or numerical, could already be in place and additional data points are collected to calibrate any uncertain parameters. Therefore, the initial DoE points do not necessarily need to be taken into account as part of the data collection process. Here, these are included in the total number of points, but for the sequential data collection processes, the initial DoE could just be thought of as what was used to train a stochastic model.

in predicting the models and the probability of failure obtained by all methods over all 20 realizations. The intent of this study is to demonstrate the feasibility of *TIGER* for calibration of accurate predictive models that can identify events of interest.

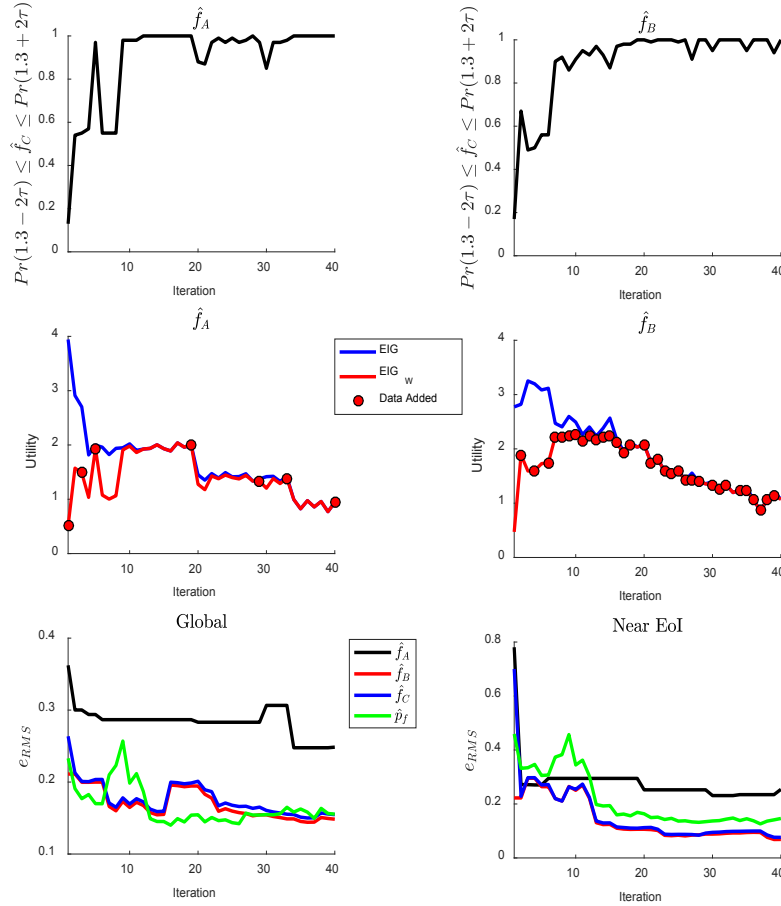
### 1. Expected information gain and weighted expected information gain

The *EIG* criterion is expected to provide space filling points in the design space. In contrast, the *EIG<sub>W</sub>* criterion is expected to put points close to the limit state of the failure region, where  $\Pr(1.3 - 2\tau \leq \hat{f}_C \leq 1.3 + 2\tau)$  is large. Here, we examine the results of a single realization of points added to either train  $\hat{f}_A$  or  $\hat{f}_B$  according to both criteria. Figure 21 displays the *EIG* value of the optimal design for each model, along with *EIG<sub>W</sub>* of that design (i.e., the probability of being in the region around the limit state). Though this probability is not used with the *EIG* criterion, the plots are shown to illustrate that points are not chosen based on high probability, and that there is significant variation in the probabilities of the chosen designs. It was also observed that 34 of the 40 points added were used to train  $\hat{f}_B$ . After approximately 10 design iterations, Figure 21 shows that once a single point was added to train  $\hat{f}_A$ , points were added to train  $\hat{f}_B$  in consecutive iterations, before a single point was again added to train  $\hat{f}_A$ . This resulted in nearly a constant  $e_{rms}$  of  $\hat{f}_A$  both globally and near the EoI as displayed in Figure 21. It was also observed that the error in  $\hat{f}_C$  is closely tied to the error in  $\hat{f}_B$ . After approximately iteration 15, the majority of points are allocated to  $\hat{f}_B$  such that the error both globally and near the EoI is reduced.



**Figure 21.** Using the *EIG* criterion, the history of the probability of being near the EoI, *EIG*, and *EIG<sub>W</sub>* values of the data points added in each design iteration, and the global and near EoI  $e_{rms}$  of each approximation.

Figure 22 shows the  $EIG_W$  value of the optimal design for each model, along with weight of that design given by  $\Pr(1.3 - 2\tau \leq \hat{f}_C \leq 1.3 + 2\tau)$ . Additionally, the unweighted  $EIG$  values at the optimal designs are shown for comparison. As expected, compared to the  $EIG$  criterion, the  $EIG_W$  places more points at regions where the probability of being near the limit state is large. As more points are added to refine the failure region, the probability goes to 1, such that  $EIG$  and  $EIG_W$  are nearly equal in later iterations. However, for the majority of the sequential design iterations, the magnitude of the utility of  $EIG$  was greater than  $EIG_W$ , since the latter was scaled by the probability of being near the EoI. This is expected, since  $EIG_W$  is designed to select data points near the EoI at the cost of reducing overall model uncertainty. As with the  $EIG$  criterion, the majority of the points (33 of 40 points added) after design iteration 5 were allocated to train  $\hat{f}_B$ , with consecutive, subsequent iterations adding points to  $\hat{f}_B$  after adding a single point to  $\hat{f}_A$ .



**Figure 22.** Using the  $EIG_W$  criterion, the history of the probability of being near the EoI,  $EIG$ , and  $EIG_W$  values of the data points added in each design iteration, and the global and near EoI  $e_{rms}$  of each approximation.

The  $e_{rms}$  at test points shown in Figure 22 shows that the overall error given by the approximate models trained using  $EIG_W$  is smaller than those obtained by  $EIG$ . This is mainly due to the large probability of being near the EoI of points added for both  $\hat{f}_A$  and  $\hat{f}_B$ . Note that here we are only comparing a single realization with each criterion, and the stochastic nature of the problem and algorithm make it difficult to generalize results. In the conclusion of this section, we describe observations and make conclusions based on the overall trends of 20 realizations.

Figures 9-11 show the model predictions obtained with  $EIG$  and  $EIG_W$  at design iterations 8, 22, and 40 (corresponding to a total points of 18, 32, and 50, respectively) for a single realization. As shown in Figure 23, the points in iteration 8 were mostly spread along the area where  $v > 0$  with both criteria.

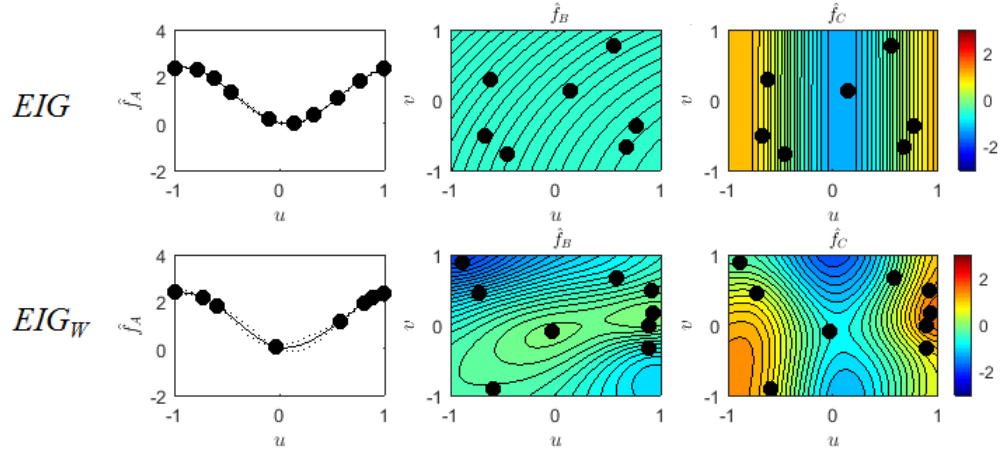


Figure 23. Model predictions after 8 design iterations (18 total points) using the *EIG* and *EIGw* criteria.

The noticeable difference in placement of points is made clear in Figs. 10 and 11. For the *EIG* criterion, by design iterations 22 and 40 the points were well-spread throughout the design space, providing accurate approximations of  $f_C$ . The *EIGw* criterion puts all points in or near the failure region.

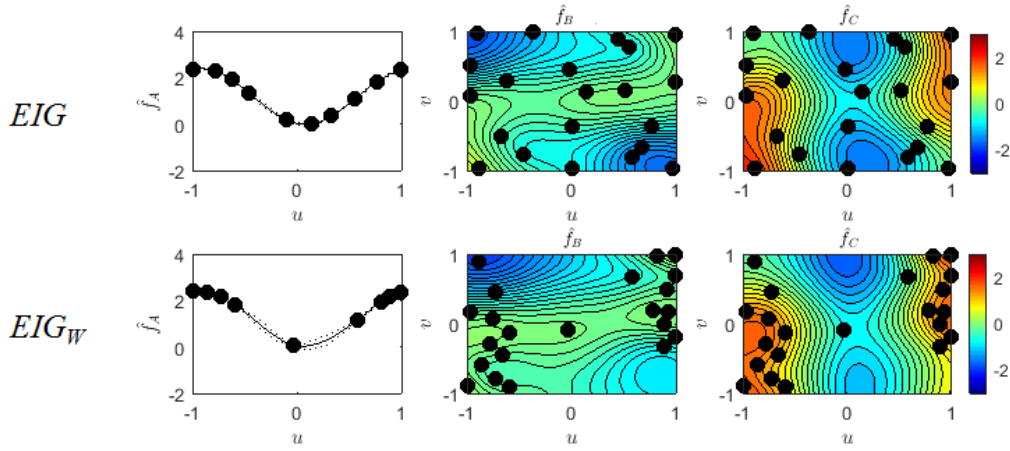


Figure 24. Model predictions after 22 design iterations (32 total points) using the *EIG* and *EIGw* criteria.

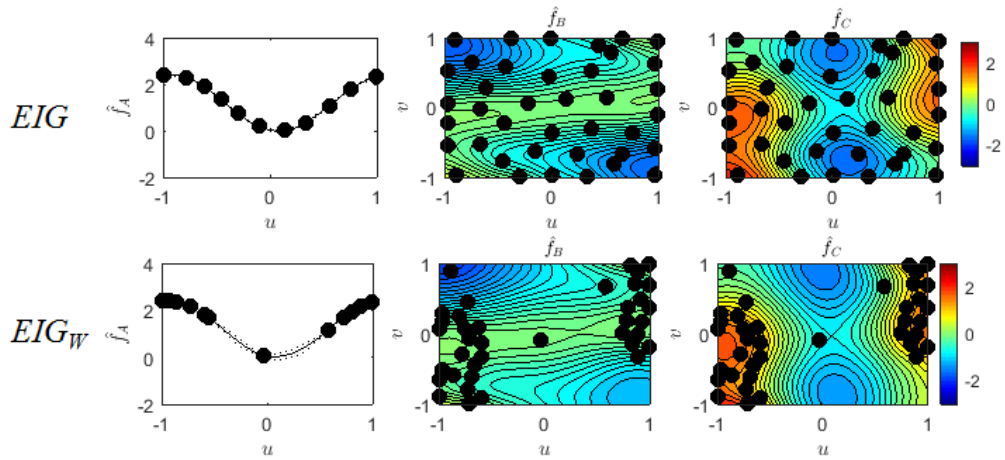


Figure 25. Model predictions after 40 design iterations (50 total points) using the *EIG* and *EIGw* criteria.

Figure 26 displays the approximation of  $p_f$  and  $\hat{f}_C \geq 1.3$ . Note that there is little difference between the approximations at design iterations 22 and 40 for the  $EIG$  criterion, and that both are close to the true  $p_f$  and failure region shown in Figure 19. By iteration 22, for  $EIG_W$ , the  $p_f$  and failure region are close to the true values, but is slightly worse than the  $EIG$  criterion.

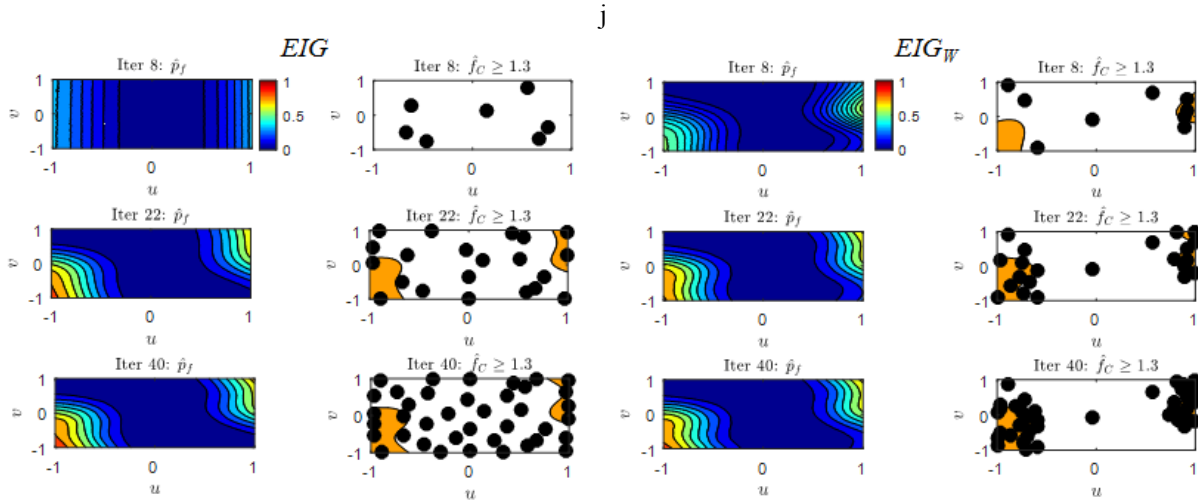
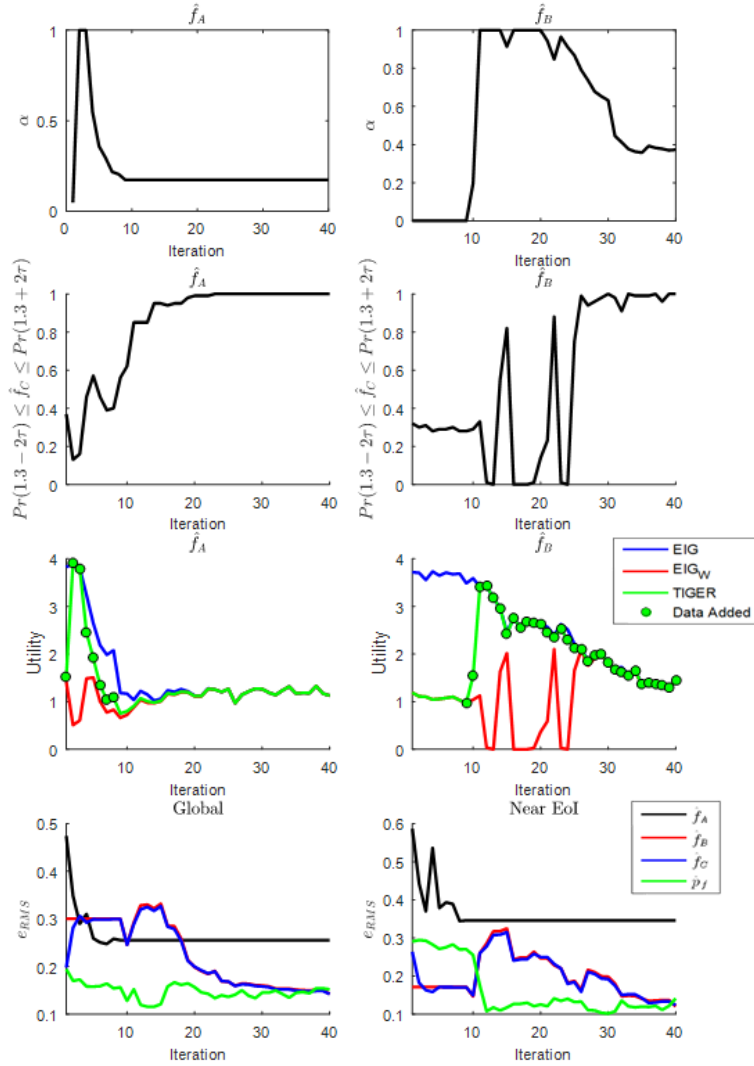


Figure 26. Predicted probability of failure and limit state boundaries after iterations 8, 22, and 40 using the  $EIG$  and  $EIG_W$  criteria.

## 2. Targeted Information Gain for Error Reduction (TIGER)

The *TIGER* criterion is expected to provide a balance between the  $EIG$  and  $EIG_W$  criteria. Figure 27 displays the weight  $\alpha$  (calculated from  $PRESS_{RMS}$  as in Figure 20) placed on each models through 40 iterations for a single realization. Initially,  $\alpha = 0$  for both  $\hat{f}_A$  and  $\hat{f}_B$  such that *TIGER* values were only functions of their respective  $EIG_W$  values. However, by the next iteration, the cross-validation error of  $\hat{f}_A$  increased  $\alpha$  to 1 such that the *TIGER* criterion was solely the  $EIG$  value. As the cross-validation error of  $\hat{f}_A$  was reduced with the addition of points globally,  $\alpha$  approached 0 through 10 iterations, thus placing more weight on  $EIG_W$ . In effect, this resulted in choosing design points with large values of  $\Pr(1.3 - 2\tau \leq \hat{f}_C \leq 1.3 + 2\tau)$  through the iterations. After iteration 10, most points were placed for  $\hat{f}_B$  with more weight on the  $EIG_W$  criterion, resulting in points with probability near 1 (i.e., near the event of interest). As with the other criteria, the number of points for  $\hat{f}_B$  (32 points) outnumbered the points for  $\hat{f}_A$  (8 points).



**Figure 27. Using the *TIGER* criterion, the history of weights  $\alpha$ , the probability of being near the EoI, *EIG* and *EIG<sub>W</sub>* values of the data points added in each design iteration, and the global and near EoI  $e_{rms}$  of each approximation.**

Figure 28 displays the model predictions for design iterations 8, 22, and 40. Unlike with the *EIG<sub>W</sub>* criterion, *TIGER* places more points globally through iterations 8 and 22. However, by 40 iterations, many points are around the failure region for a comparable approximation to what was achieved with *EIG<sub>W</sub>*.

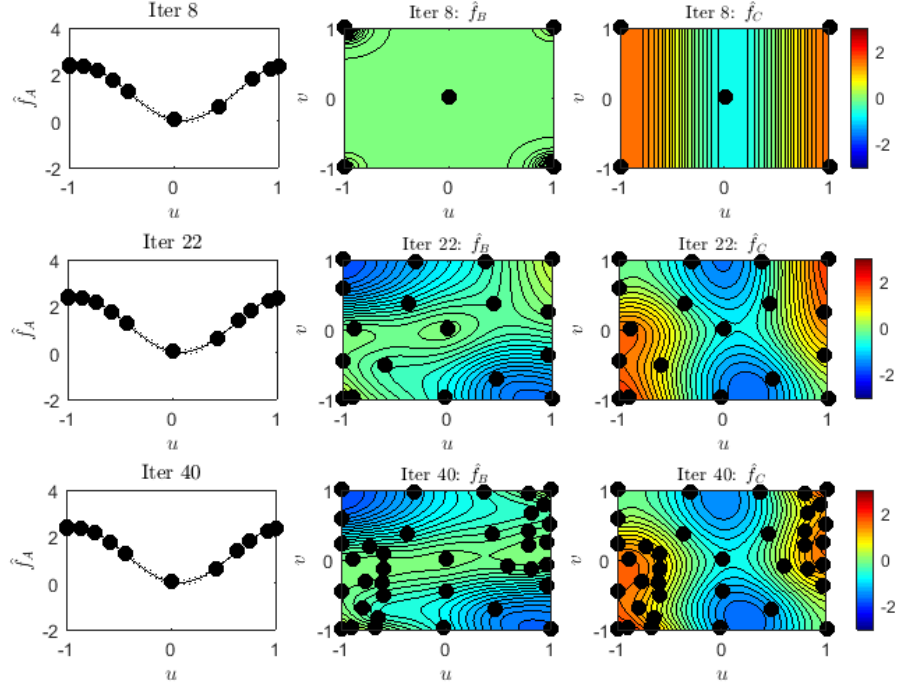


Figure 28. Model predictions after design iteration 8, 22, and 40 (18, 32, and 50 total points, respectively) using the *TIGER* criterion.

Figure 29 displays the predicted probability of failure and limit state boundaries using the *TIGER* criterion. After 8 design iterations, the failure regions along  $u = \pm 1$  are found, and are refined by iteration 22. After 40 iterations, the model accurately found the probability of failure and limit state boundaries.

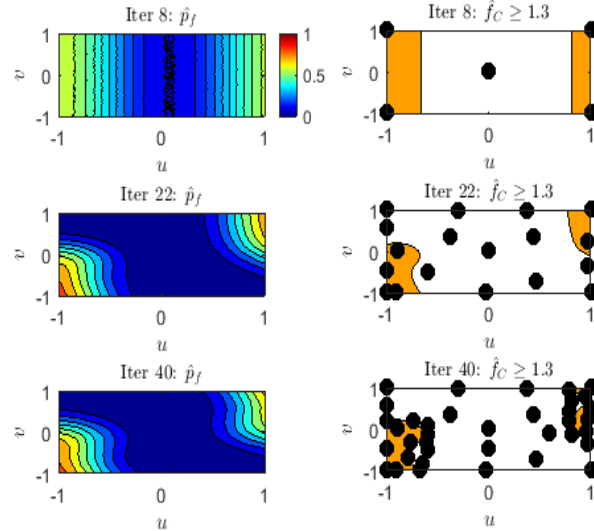


Figure 29. Predicted probability of failure and limit state boundaries after design iterations 8, 22, and 40 (18, 32, and 50 total points, respectively) using the *TIGER* criterion.

### C. Comparison of Experimental Designs

One advantage of the information theory based sequential data collection methods are that they can determine both the optimal data point and experiment by comparing the information gain from each source. The single realization results showed that the majority of designs were used to train  $\hat{f}_B$ , which is more complex than the one-dimensional quadratic  $f_A$ . Over 20 realizations of each method, it was observed that each method allocated points to train  $\hat{f}_B$  nearly



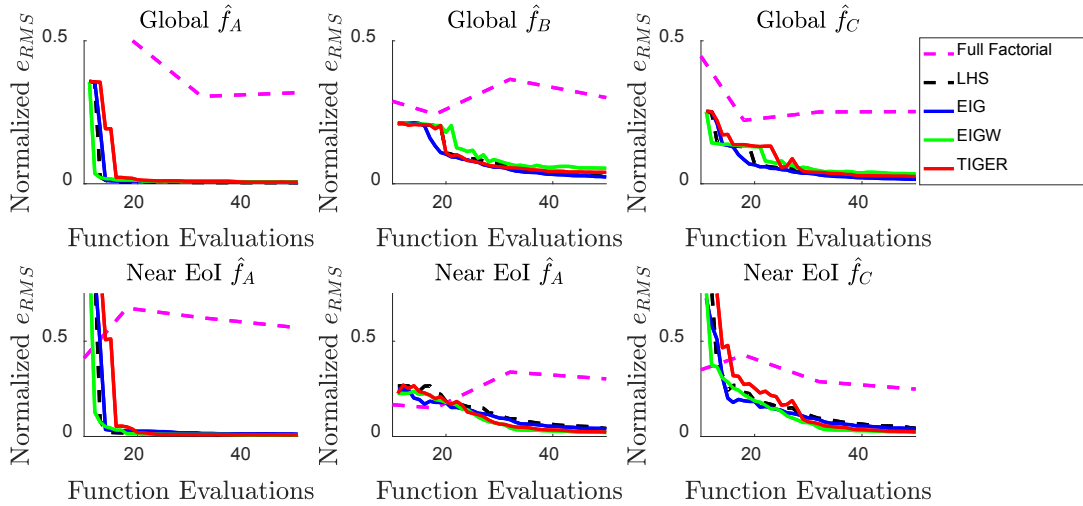
80% of the time, as shown in Table 7. Note that this study does not account for model complexity or cost of the experiment, but measurement uncertainty is taken into account in the expected information gain.<sup>4</sup>

**Table 7. Median percentage of points used to train  $\hat{f}_B$  with each sequential data collection criterion**

Criterion	Median $\hat{f}_B$ training points (% of total number of points)
<i>EIG</i>	80
<i>EIG<sub>W</sub></i>	80
<i>TIGER</i>	78

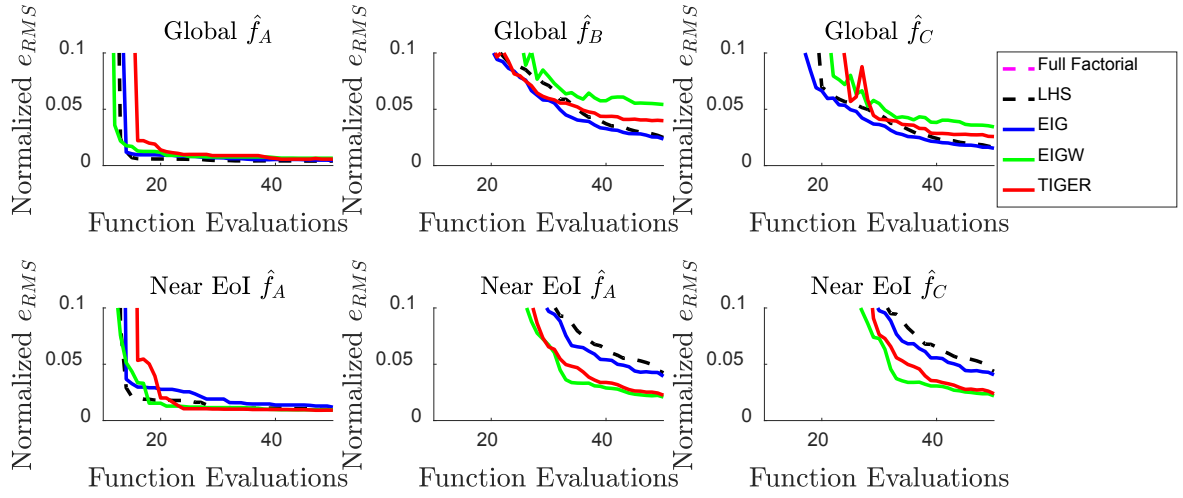
To compare the data collection methods, the root mean square error ( $e_{rms}$ ) relative to the nominal function was calculated at 100 random points in the design space. Additionally, the accuracy near the EoI (corresponding to the area  $\pm 2\tau$  from the limit state) was also calculated at 100 random points. Figure 30 shows a comparison of these errors normalized by the range of the true values of  $f_C$  for the  $e_{rms}$  values, and Figure 31 zooms in on the normalized  $e_{rms}$  axis. The comparisons were made for the median of 20 realizations of each case. Additionally, the comparison was made to design of experiments with LHS and full factorial (FF) design of experiments.

For global and near EoI accuracy, the FF DoE had the worst performance. Of the sequential data collection methods, the *EIG* criterion provided global accuracy, but was outperformed near the EoI by *EIG<sub>W</sub>* and *TIGER*, which are both driven to put points near the EoI. As expected, *EIG<sub>W</sub>* resulted in the smallest overall error near the EoI, but it did not perform as well globally as *EIG* and *TIGER*. Finally, *TIGER* provided a balance in accuracy between *EIG* and *EIG<sub>W</sub>*, performing well both globally and near the EoI. For 32 and 50 total function evaluations, it performs as well as *EIG<sub>W</sub>* near the EoI, which can be attributed to the *TIGER* criterion placing more weight on designs close to the EoI as the accuracy of the model predictions increased. Note that the level of ‘aggression’ in locating the EoI can be tuned by the user-defined form of  $\alpha(PRESS_{RMS})$ .



**Figure 30. Comparison of  $e_{rms}$  of models at test points globally and near the EoI for FF, LHS, EIG, EIG<sub>W</sub>, and TIGER criteria.**

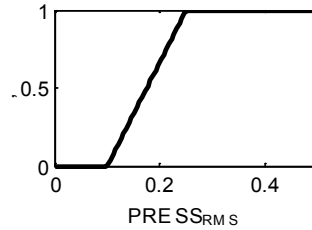
<sup>4</sup> That is, the uncertainty in the model could be large, but a large measurement uncertainty would result in less information gained from data collected in the experiment.



**Figure 31. Zoom in on comparison of  $e_{rms}$  of models at test points globally and near the EoI for FF, LHS, EIG, EIGW, and TIGER criteria.**

#### D. Adjusting the Level of TIGER ‘Aggression’

The form of  $\alpha$  as a function of  $PRESS_{RMS}$  can be adjusted to change the level of ‘aggression’ of the search for the EoI. That is,  $\alpha$  can be changed to either increase or decrease the emphasis put on EIG or EIGW. This scenario might occur when an analyst has some idea of how accurate a model can be, so a lower bound is set for  $PRESS_{RMS}$ . We examined  $\alpha$  bounded by  $[0, 1]$  and linear between  $PRESS_{RMS}$  values between 0.1 and 0.25, as shown in Figure 32. Note that this means that for  $PRESS_{RMS}$  values less than 0.1, the objective function is only EIGW.



**Figure 32. More aggressive  $\alpha$  as a function of  $PRESS_{RMS}$**

Figure 33 displays the model predictions using the more aggressive  $\alpha$ . Comparing this results to Figure 28, it was obvious that the more aggressive  $\alpha$  places the majority of points near the EoI.

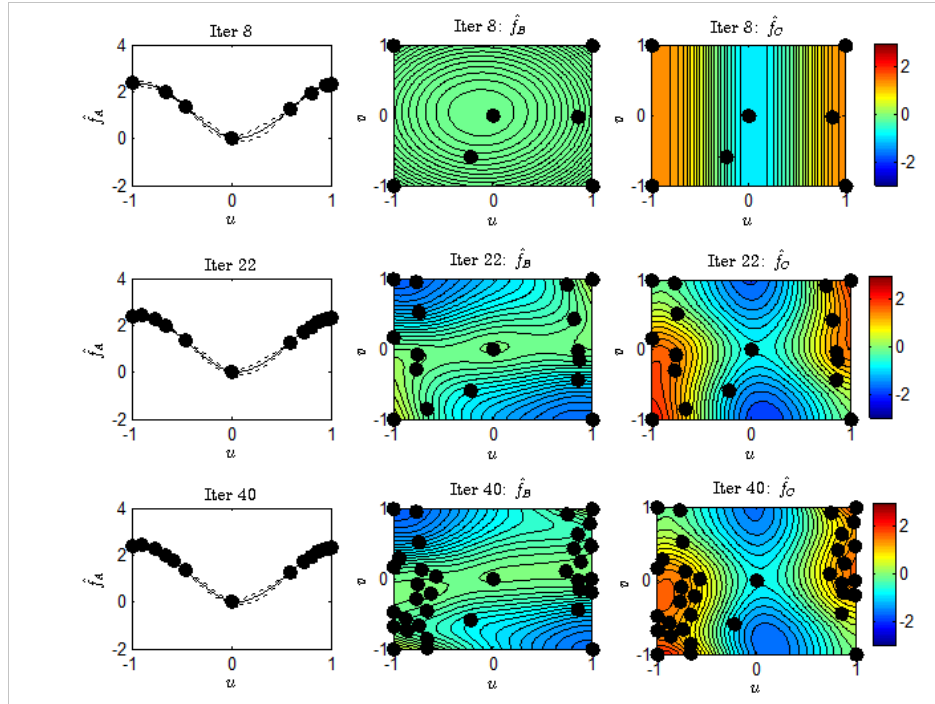


Figure 33. Model predictions after design iteration 8, 22, and 40 (18, 32, and 50 total points, respectively) using the *TIGER* criterion.

Figure 34 displays the predicted probability of failure and limit state boundaries using the *TIGER* criterion. After 8 design iterations, the failure regions along  $u = \pm 1$  are found, and are refined by iteration 22. After 40 iterations, the model accurately found the probability of failure and limit state boundaries.

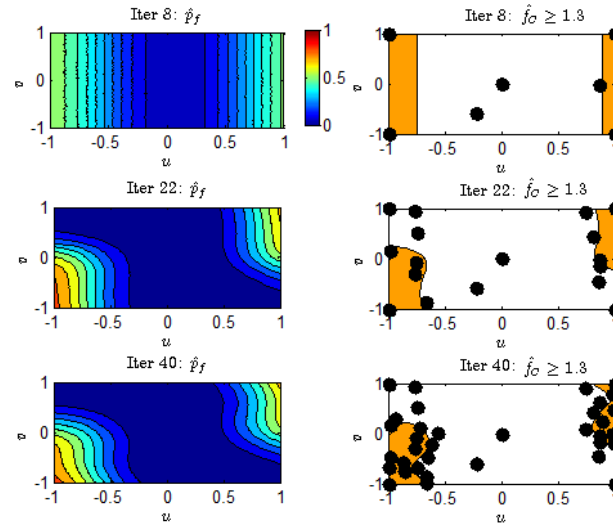


Figure 34. Predicted probability of failure and limit state boundaries after design iterations 8, 22, and 40 (18, 32, and 50 total points, respectively) using the *TIGER* criterion.

### III. Application Problem: Flutter Boundary and Critical Limit Cycle Oscillation Amplitude Identification

Hypersonic aircraft structures are subjected to intense, coupled, fluid-thermal-structural loading during high-speed flight. The aeroelastic model components of the fluid-structure interaction are shown in Figure 35. Considering a panel in hypersonic flow, the flow acting on the structure results in aerodynamic pressure on the wetted surface of the panel. This leads to elastic deformation of the panel into the flow field, resulting in feedback on the flow.

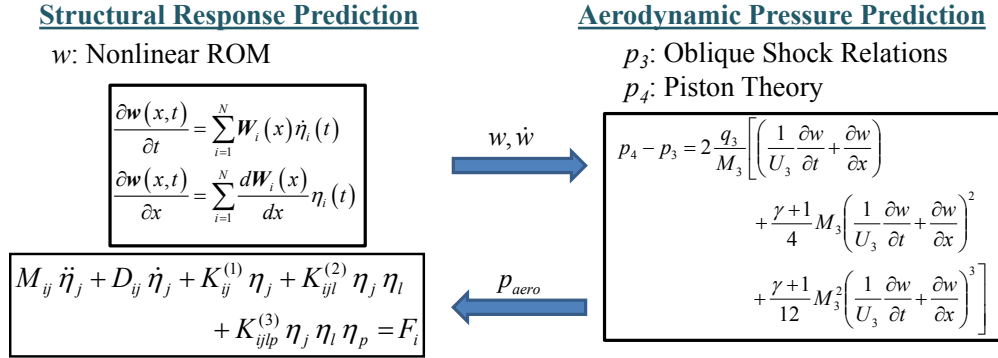


Figure 35. Aeroelastic coupling and aeroelastic solution, where panel slope and velocity is transferred to piston theory and aerodynamic pressure is transferred to the structural solution

The aeroelastic model was used to investigate the impact of model uncertainty on nonlinear panel flutter. Flutter is an aeroelastic instability where the amplitude of vibration of a structural component in a flow field increases without a bound. In the case of a panel, nonlinear membrane stretching provides a stabilizing effect that restrains the panel motion to a bounded amplitude for limit cycle oscillations (LCO). Panel flutter not only provides an extreme response scenario for this coupled system, but is a design constraint of aerospace structures. Unsurprisingly, high fidelity analyses are computationally expensive, such that researchers have examined efficient methods of coupling the required aerodynamic and structural analyses in the time domain.<sup>32</sup>

Perez et al.<sup>33</sup> predicted frequencies and amplitudes associated with limit cycle oscillations by coupling a structural reduced order model (ROM) with aerodynamic pressure predictions from piston theory. In their work, the model-form error in piston theory was identified and used to correct the aerodynamic pressure predictions. The ROM built with 6 linear modes was constructed for an isotropic 2-D panel clamped along the sides perpendicular to the direction of the flow. The geometric, material properties of the panel, and flow conditions are shown in Table 8. The ROM was constructed with an in-house FEA beam model based on a co-rotational formulation capable of analyzing problems with large rotations and small strains. The FEA model was built using 40 beam elements, a total of 123 degrees-of-freedom. Figure 35 also provides a schematic of the solution of the aeroelastic problem.

Table 8. Aeroelastic model parameters

Parameter	Value	CV (%)	Unit
Mach Number	5 - 12		
Altitude	30	--	km
Freebody Surface Inclination	5	--	deg
Panel Length	1.5	--	m
Panel Thickness	2	--	mm
Density	4539	--	kg/m <sup>3</sup>
Modulus of Elasticity	113	1	GPa
Temperature	226	1	K
Pressure	1.2	1	kPa

The panel displacement and velocity,  $w(x,t)$  and  $\dot{w}(x,t)$ , are computed using the structural ROM. The panel deformation serves as a boundary condition to the flow problem, for which oblique shock relations are used to compute the pressure after the shock, and 3rd-order piston theory to obtain the pressure at the deformed surface of

the panel.

Two limit state conditions were considered as the EoIs for this study. The first one was the flutter boundary of the panel. Typically, the panel stiffness has to be increased in order to avoid reaching its flutter boundary for pre-defined flow conditions. The other limit state considered was a critical limit cycle oscillation amplitude. In practice, the definition of this critical amplitude would have to take into consideration the fatigue life of the panel as well as the aerodynamic performance.

The identification of these limit states is done as follows. For a pre-defined altitude, the Mach number of the flow is initialized and the type of limit state of interest specified (i.e., flutter boundary or critical limit cycle oscillation amplitude). Then, the dynamic response of the panel to a small initial displacement is computed by marching the structural equations of motion in time using a Newmark- $\beta$  algorithm. This algorithm is unconditionally stable for linear problems. For nonlinear problems, as the present case, the time step used in the integration of the equations of motion is chosen as the largest one that leads to a converged solution. After marching the solution for a pre-defined length of time the integration stops and the structural response is used to determine if the conditions for the limit state selected are met. The existence of a limit cycle oscillation (i.e., amplitude  $a_{LCO} > 0$ ) indicates that the flutter boundary has been crossed. If the amplitude of the limit cycle oscillation is larger than the critical one means that the second limit state has been reached. The process is stopped if the limit state condition is reached; otherwise, the Mach number of the flow is increased. This process is repeated until the condition for the limit state chosen is met.

Even with the use of surrogate CFD models and ROMs, the identification of the limit states can be computationally expensive due to the stepping of Mach numbers combined with the Newmark- $\beta$  algorithm. In the presence of uncertainty, particularly when we seek to propagate uncertainty through Monte Carlo sampling, the analysis becomes even more computationally demanding. Therefore, in this study, we sought to use information gain techniques to identify the limit states with classifiers that consider the amplitude as a function of Mach number in the presence of input uncertainties. The input temperature, pressure, and elastic modulus were considered to be uncertain, with known coefficient of variation shown in Table 8. To identify the limit states, we used a Gaussian Naïve Bayes classifier, which classifies a point based on Mach number as a no flutter, non-critical LCO, or critical LCO point. Additionally, the probability of belonging to each class is provided. For detailed information about the Naïve Bayes classifier, the reader is referred to the Appendix.

#### E. Naïve Bayes Classifiers for Flutter and Critical LCO Amplitude

The Naïve Bayes (NB) classifier, described in detail in the Appendix, was used to find the probability of flutter and probability of exceeding the critical LCO amplitude  $a_{LCO} = 1$  unit thickness at a given Mach number. Temperature, pressure, and elastic modulus were considered as random variables with known aleatory uncertainty, but Mach number was the only the experimental design variable. Therefore, no assumptions about conditional independence among design variables were necessary.

To find the probability of flutter ( $a_{LCO} > 0$ ) and critical LCO for ( $a_{LCO} \geq 1$ ), two NB classifiers with two classes each with binary labels  $y$  were defined:

- 1) Flutter:  $y_F = 0$  for  $a_{LCO} = 0$  (no flutter) and  $y_F = 1$  for  $a_{LCO} > 0$  &  $a_{LCO} < 1$  (non-critical LCO)
- 2) Critical LCO:  $y_{crit} = 0$  for  $a_{LCO} < 1$  (no flutter or non-critical LCO) and  $y = 1$  for  $a_{LCO} \geq 1$  (critical LCO).

Note that an alternative approach could define a single NB classifier with three classes for no flutter, non-critical LCO, and critical LCO.

Since Mach number was the only attribute and for each two class model, there are a total of four uncertain parameters for each model, which are the means and standard deviations of each Gaussian distribution for each class. Therefore, in the context of the notation of this study, the Mach numbers serve as the design variables  $x$ , while the means and standard deviations of each classifier are the uncertain parameters  $\theta$  which are calibrated with test data. The priors for the classification problem (i.e.,  $\Pr(y = y_k)$  for class  $k$ , refer to Eq. (38) in the Appendix) were set as uniform, rather than the empirical probability for the data. The NB were trained using an initial DoE of nine points, which were randomly sampled from the  $M \sim U(5, 12)$ , with three points belonging to each class.

#### F. Data Collection Using TIGER

Since there were only three possible outcomes possible at each data point (i.e., no flutter, non-critical LCO, and critical LCO),  $EIG$ ,  $EIG_W$ , and  $U_{TIGER}$  could be calculated explicitly at each  $M$ . This was achieved by simply re-fitting the NB classifier for each possible class at each candidate  $M$ . The expected information gain was then calculated with Eq. (32) for  $K = 2$  classes, which is provided explicitly for this problem.

$$EIG(M) = \frac{1}{K^2} \sum_{j=1}^K \sum_{i=1}^K \ln \left[ \frac{\Pr(y = y_i | \theta_{y=y_j})}{\Pr(y = y_i | \theta)} \right] \Pr(y = y_i | \theta_{y=y_j}) \quad (32)$$

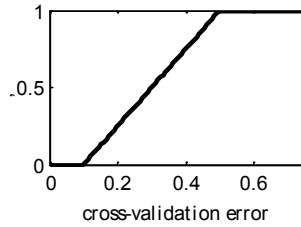
In this notation,  $\Pr(y = y_i | \theta)$  represents the probability of the  $i^{\text{th}}$  class given the prior  $\theta$ , and  $\Pr(y = y_i | \theta_{y=y_j})$  is the re-fitted NB classifier given  $y = y_j$ .

For  $EIG_W$ , the event of interest was determined by the class. Since the focus of this study was identifying flutter ( $y_F = 0$ ) and the onset of critical amplitude ( $y_{crit} = 1$ ), the EoIs for classes 0 and 1 are the boundaries between the classes for each model, respectively.

$$EIG_W |_{y_F=0} (M) = EIG(M) [\Pr(y_F = 0 | \theta) \Pr(y_F = 1 | \theta)] \quad (33)$$

$$EIG_W |_{y_{crit}=1} (M) = EIG(M) [\Pr(y_{crit} = 0 | \theta) \Pr(y_{crit} = 1 | \theta)] \quad (34)$$

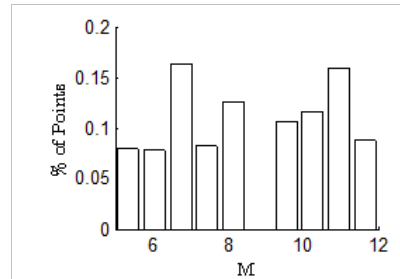
The optimization problem shown in Eq. (25) could then be solved to find the optimum  $M$  for a given EoI, where the candidate Mach numbers were sampled from a discrete set of 101 uniformly spaced points on  $M = [5, 12]$ . In this study, the goal was to find the designs with the maximum utilities for re-training each classifier. Therefore, Eq. (25) was solved to find the optimum  $M^*$  for each EoI for a given class using Eqs. (32)-(34), for a total of two new points in each iteration. The weight  $\alpha$  was again a function of the cross-validation error and the relationship is shown in Figure 36. In this example, the cross-validation error is the misclassification rate calculated from the number of misclassifications between the true label and NB classifier in a leave-one-out fashion.



**Figure 36. Value of  $\alpha$  as a function of cross-validation error**

The algorithm using the maximum *TIGER* criterion was run for 10 design iterations for 20 realizations. Since two points were added to train each classifier in each design iteration, there were a total of 29 design points for each realization, including the initial DoE of nine points. For each optimum  $M^*$  found using *TIGER*, a random sample of the input uncertainties (i.e., temperature, pressure, and elastic modulus) were obtained and the corresponding  $a_{LCO}$  was calculated using the structural ROM and CFD surrogate. The points were then added to the data set and available to both classifiers in the next design iteration.

As a point of comparison to a NB classifier with a large number of points, 1,560 points were obtained by sampling  $M$  from 5 to 12 with random temperature, pressure, and elastic modulus. These 1,560 points were used to fit ‘true’ NB classifiers. Figure 37 displays a histogram of the Mach numbers of the 1,560 points. Based on the histogram, a slightly larger percentage of the sampled points were around each EoI, and a small region near  $M = 9$  was unsampled.



**Figure 37. Histogram of the Mach number of the bank of 1,560 points.**

Figure 38 shows the 1,560 points with the original label and the labels given by the ‘true’ NB classifiers. It was observed that there is some misclassification of points at the limit states of the EoIs.

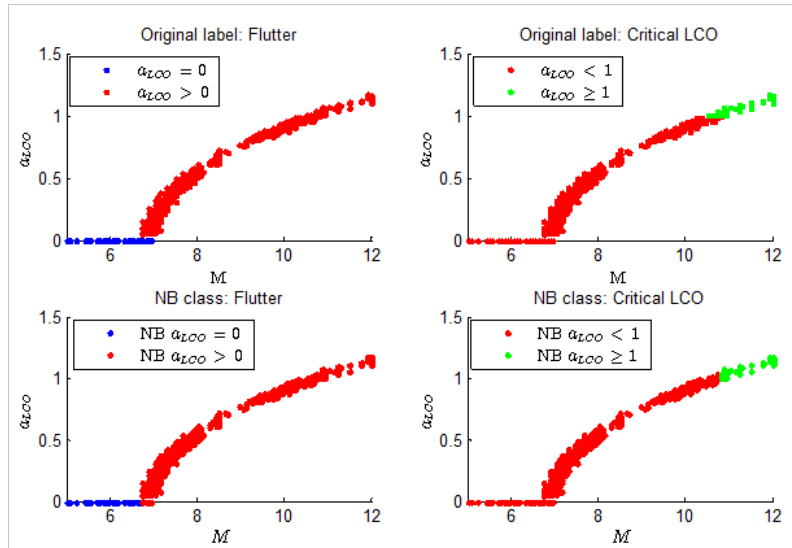


Figure 38. Original classes of 1,560 points and classes given by ‘true’ NB classifiers.

Using the ‘true’ classifiers, the posterior probability of 1,000 test points uniformly spaced on  $M = [5, 12]$  with nominal values for uncertain inputs was calculated and is shown in Figure 39. These ‘true’ NB classifiers were used to make comparisons to the classifiers obtained with *TIGER* at the 1,000 test points.

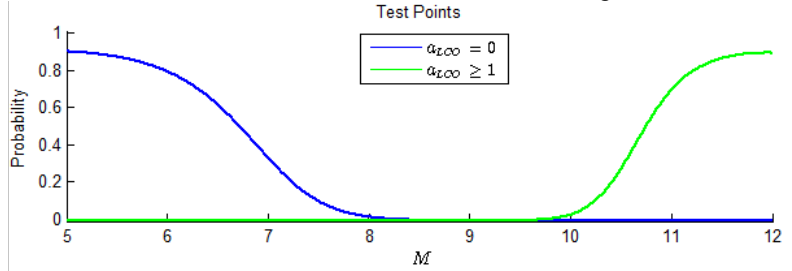


Figure 39. Posterior probabilities of 1,000 test points from  $M$  from 5 to 12 for the ‘true’ NB classifiers.

### G. Flutter and Critical LCO Identification Results

The distribution of optimal  $M$  for each classifier using *TIGER* over 20 realizations of 10 design iterations is displayed in Figure 40. Figure 40 also displays a histogram that shows the majority of points were near both EoIs, namely  $M = 7$  for flutter and  $M = 10$  for critical LCO.

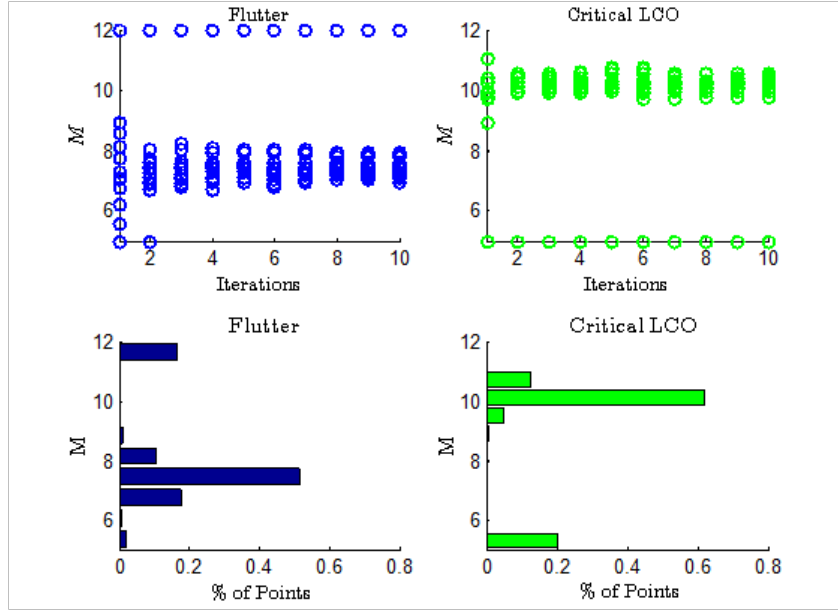


Figure 40. Optimal  $M$  added using *TIGER* over 10 iterations and 20 realizations for flutter and critical LCO classifiers with the histogram representing the frequency of each optimal  $M$ .

For each classifier, only 20% of points were placed away from the EoI. This is due to  $\alpha$  being equal to zero approximately 80% of the time, which puts all of the weight on  $EIG_W$  (targeting the EoIs), as shown in Figure 41. The small percentage of points for  $\alpha > 0$  resulted in the points away from the EoIs.

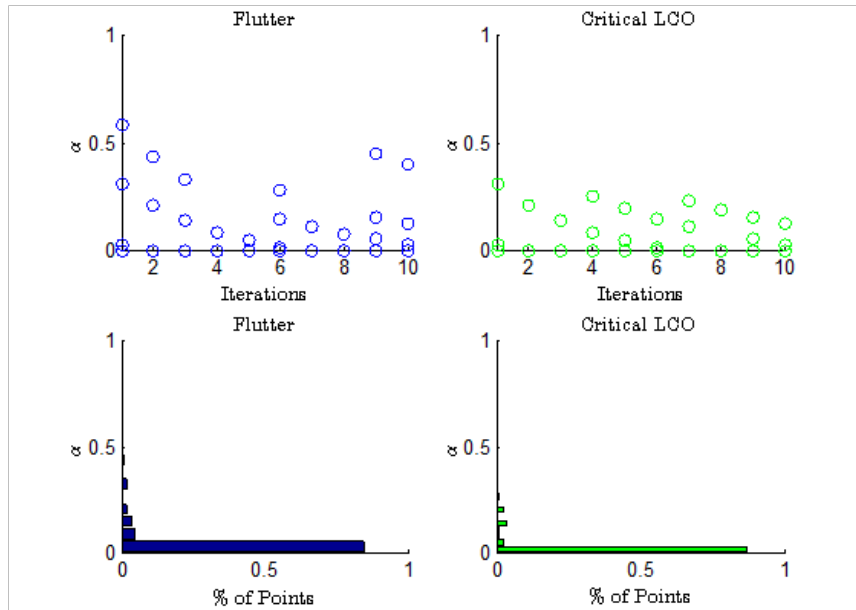
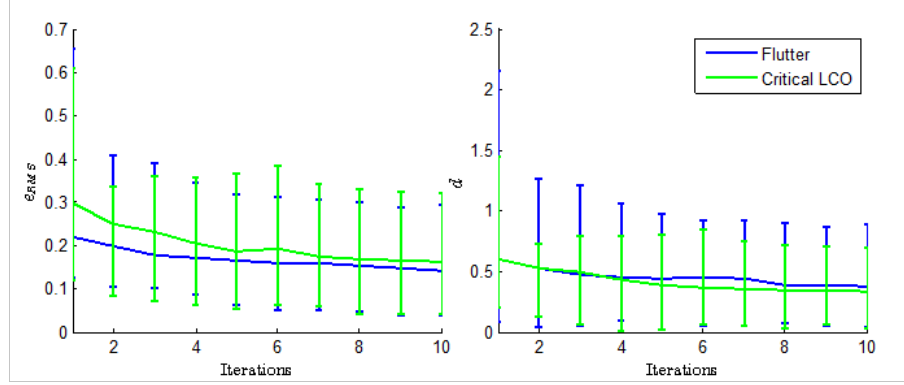


Figure 41. Weight  $\alpha$  for *TIGER* over 10 iterations and 20 realizations for flutter and critical LCO classifiers with the histogram representing the frequency of each  $\alpha$ .

Figure 42 displays the  $e_{RMS}$  and area metric of the posterior probabilities given by the NB classifier compared to the ‘true’ posterior probabilities over 1,000 test points uniformly spaced on  $M = [5, 12]$ . The  $e_{rms}$  was calculated for  $M \leq 9$  for the flutter classifier and for  $M \geq 9$  for the critical LCO classifier. It was observed that the error in each probability reduced as more points were added through design iterations, with both EoIs predicted with comparable accuracy by the NB classifiers after 10 iterations.





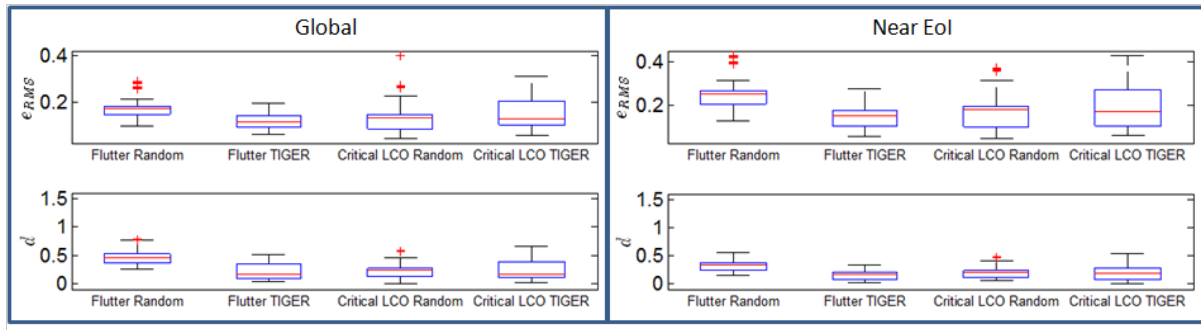
**Figure 42.**  $e_{rms}$  and area metric  $d$  given by ‘true’ posterior probabilities and those given by NB classifier. The solid line represents the median with the 5<sup>th</sup> and 95<sup>th</sup> percentiles represented by error bars.

Additionally, the difference in area between the ‘true’ and NB classifier probability curves was measured over the same 1,000 test points of  $M$ . Equation (35) displays the area metric<sup>34</sup>  $d$ , which is a commonly used measure for validation of computational models, where  $y_i$  represents  $y_F$  or  $y_{crit}$ . As with the  $e_{rms}$ , the domain of  $M$  was used to calculate  $d$  for  $M \leq 9$ , and for  $M \geq 9$  for the critical LCO classifier.

$$d = \int_M |\Pr(y = y_i)_{true} - \Pr(y = y_i | \theta)| dM \quad (35)$$

The area metric results shown in Figure 42 mostly agree with the  $e_{rms}$ , where the critical LCO classifier showed relatively the same improvement as the flutter classifier.

As another assessment of the *TIGER* classifiers, 20 realizations of random DoEs of 29 points were created for comparison with *TIGER* for an equal number of training points. For these random DoEs, the Mach number was sampled randomly from the 1,560 points that were used to make the ‘true’ classifiers. Figure 43 compares the global and near EoI  $e_{rms}$  over 1,000 test points for a total of 29 points obtained through *TIGER* and by 20 realizations of a randomly sampled DoE of 29 points. That is, the *TIGER* results shown are obtained from the final classifier trained by after 10 design iterations. For the region near EoI, consider the ranges of  $6.5 \leq M \leq 8$  for flutter and  $9.5 \leq M \leq 11$  for critical LCO. For flutter, *TIGER* outperforms the random DoE based on  $e_{rms}$  and the area metric. The LCO classification is more competitive for both global and near the EoI. This success can be attributed to the placement of points near the limit state of the EoI by *TIGER*, whereas the random DoEs were more uniformly distributed but with slightly larger numbers of samples near the EoIs, as shown previously in Figure 37.



**Figure 43.** Boxplot of  $e_{rms}$  over 1,000 test points for 29 total points over 20 realizations.

Based on the results of this application problem, *TIGER* is slightly better than using a random DoE for the same number of total points. However, *TIGER* may be useful if a large number of points already exist in the DoE. For example, the random DoE had larger error in estimating the probability of flutter. As the *TIGER* criterion places many points near the EoI limit state, it could be useful in reducing the error of the random DoE. Additionally, the user-defined form of  $\alpha$  as a function of the cross-validation error can be adjusted. As observed over 20 realizations,  $\alpha = 0$

for the majority of iterations. Changing the  $\alpha$  function could result in more global, space-filling points that increase the accuracy of the NB classifiers trained with data from the *TIGER* criterion.

#### IV. Summary

A data collection approach based on expected information gain was developed to determine the optimal design of experiments to build models that accurately predict targeted events while maintaining global accuracy. A bi-objective Targeted Information Gain for Error Reduction (*TIGER*) approach was formulated to balance targeting the event with global accuracy. The first investigation focused on the accurate prediction of the limit state and probability of failure of a two-dimensional analytical function. A sequential data collection approach was taken to locate the design with the maximum *TIGER* criterion, and noisy test data was obtained to calibrate Gaussian process surrogates until the maximum number of allowable tests was reached. Additionally, it allowed the comparison of two information sources, such that it was possible to allocate experiments for training data between the models. It was observed that *TIGER* is an effective approach based on the comparison of prediction errors with three other data collection methods.

Sequential data collection with the *TIGER* criterion was used to find the optimal data points for training a classifier for flutter and critical LCO. This method was shown to be much more accurate than a random design of experiments for the same number of points, and it displayed the ability to put points near the boundary, which is useful when working with accurate global models.

Finally, both examples in this study used statistical models (i.e., naïve Bayes classifiers and Gaussian process surrogates). Therefore, the method was dependent on the initial design of experiments and heuristics (e.g., the weight parameter  $\alpha$  based on cross-validation error). For physical or numerical models, where initial training points are unnecessary, such parameters may not be applicable, so the user would have to define these parameters or additional heuristics for their particular problem. For example, without an initial set of training points, cross-validation error has no meaning, which leads to an undefined value of  $\alpha$ . To implement the *TIGER* methodology, the user could define an additional heuristic, such as using the expected information gain criterion to select points for the first few design iterations before switching to the *TIGER* formulation.

Future work includes investigating the implementation of an experimental budget given the cost of tests. In the modified camelback example presented, the majority of points were added to train the more complex two-dimensional model rather than the one-dimensional model. The addition of cost would play an interesting role in the allocation of experiments among multiple models.

#### Appendix: Gaussian Naïve Bayes Classifiers

The Naïve Bayes (NB) algorithm<sup>35,36</sup> is used to estimate discrete classifications  $y$  for attributes  $x$ , which can be considered as design variables in this study. For NB classifiers, an important assumption is that the  $n$  attributes or design variables ( $x_i$  for the  $i^{\text{th}}$  attribute) are conditionally independent of each other for a given  $y$  as given by Eq. (36)

$$\Pr(x_1, \dots, x_n | y) = \prod_i^n \Pr(x_i | y) \quad (36)$$

For  $K$  classes and  $n$  attributes, Bayes' rule is used to describe the probability that  $y$  will take on its  $k^{\text{th}}$  possible value as shown in Eq. (37), given the assumptions of conditional independence of the  $x_i$ .

$$\Pr(y = y_k | x) = \frac{\Pr(y = y_k) \prod_i \Pr(x_i | y = y_k)}{\sum_j \Pr(y = y_j) \prod_i \Pr(x_i | y = y_j)} \quad (37)$$

At a new  $x$ ,  $x^{\text{new}} = \langle x_1, \dots, x_n \rangle$ , the probability that it belongs to a class can be calculated by Eq. (37). To classify  $x^{\text{new}}$  to the most probable value of  $y$ , the NB classification rule is

$$y \leftarrow \arg \max_{y_k} \frac{\Pr(y = y_k) \prod_i \Pr(x_i | y = y_k)}{\sum_j \Pr(y = y_j) \prod_i \Pr(x_i | y = y_j)} \quad (38)$$

which simplifies to

$$y \leftarrow \arg \max_{y_k} \Pr(y = y_k) \prod_i \Pr(x_i | y = y_k) \quad (39)$$

because the denominator is independent of  $y_k$ .

For continuous inputs  $x_i$ , as considered in this study, we can represent the distributions  $\Pr(x_i | y)$  by assuming that for each possible discrete class  $y_k$ , the distribution of each  $x_i$  is Gaussian. To train the NB classifier, the mean  $\mu_{ik}$  and standard deviation  $\sigma_{ik}$  of each Gaussian is estimated, along with the prior  $\Pr(y = y_k)$ .

$$\mu_{ik} = E[x_i | y = y_k] \quad (40)$$

$$\sigma_{ik}^2 = E[(x_i - \mu_{ik})^2 | y = y_k] \quad (41)$$

Therefore, for a Gaussian NB classifier, there are  $2nK$  uncertain parameters to estimate independently, which make up the uncertain parameters  $\theta$ . Maximum likelihood estimators are used to estimate the parameters as shown in Eqs. (42) and (43),

$$\hat{\mu}_{ik} = \frac{1}{\sum_j \delta(y^{(j)} = y_k)} \sum_j X_i^j \delta(y^{(j)} = y_k) \quad (42)$$

$$\hat{\sigma}_{ik}^2 = \frac{1}{\sum_j \delta(y^{(j)} = y_k)} \sum_j (x_i^{(j)} - \hat{\mu}_{ik})^2 \delta(y^{(j)} = y_k) \quad (43)$$

where  $\delta(y = y_k)$  is 1 if  $Y = y_k$  and 0 otherwise, such that the role of  $\delta$  is to select the training examples for which  $y = y_k$ .

### Acknowledgments

This research is sponsored by the Air Force Office of Scientific Research (AFOSR). The authors thank Adam Culler and Erin DeCarlo for the MATLAB codes for the aerothermal model and Bayesian calibration.

### References

- <sup>1</sup>Glass, C.E., and Hunt, L.R., "Aerothermal Tests of Spherical Dome Protuberances on a Flat Plate at a Mach Number of 6.5," NASA TP-2631, 1986.
- <sup>2</sup>Lindley, D. V., "On a measure of the information provided by an experiment," *The Annals of Mathematical Statistics*, pp. 986-1005, 1956.
- <sup>3</sup>Huan, X., and Marzouk, Y.M., "Simulation-based optimal Bayesian experimental design for nonlinear systems," *Journal of Computational Physics*, Vol. 232, No. 1, 288-317, 2013.
- <sup>4</sup>Liepe, J., Filippi, S., Komorowski, M., and Stumpf, M. P., "Maximizing the information content of experiments in systems biology," *PLoS computational biology*, Vol. 9, No. 1, 2013.
- <sup>5</sup>Sebastiani, P., Wynn, H. P., "Maximum entropy sampling and optimal Bayesian experimental design," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 62, No. 1, 2000, pp. 145-157
- <sup>6</sup>Bryant, C., Terejanu, G., "An Information-Theoretic Approach to Optimally Calibrate Approximate Models," *Proc., The 50th AIAA Aerospace Sciences Meeting*, Nashville, Tennessee, 2012
- <sup>7</sup>Haldar, A., and Mahadevan, S., *Probability, Reliability and Statistical Methods in Engineering Design*, John Wiley & Sons, New York, 2000.
- <sup>8</sup>Smarslok, B. P., Culler, A. J., and Mahadevan, S., "Error Quantification and Confidence Assessment of Aerothermal Model Predictions for Hypersonic Aircraft," *Proc., 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conf.*, AIAA 2012-1965, Honolulu, HI, 2012.
- <sup>9</sup>DeCarlo, E.C., Mahadevan, S., and Smarslok, B.P., "Bayesian Calibration of Aerothermal Models for Hypersonic Air Vehicles," *Proc., 15th AIAA Non-Deterministic Approaches Conf.*, AIAA 2013-1683, Boston, MA, 2013.

- <sup>10</sup>Balch, M.S. and Smarslok, B.P. (2014). "A Pre-Validation Study on Supersonic Wind Tunnel Data Collected from Legacy Aerothermal Experiments." Proc. AIAA SciTech Conference, AIAA 2014-0812, National Harbor, MD.
- <sup>11</sup>Ostoich, C., Bodony, D.J., Geubelle, P.H., "Development and Validation of a First Principles Fluid-Thermal Multi-Physics Solver for Hypersonic Boundary Layer Heat Transfer Problems," *Proc., 52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conf.*, AIAA 2011-1964, Denver, CO, 2011.
- <sup>12</sup>Ashley, H., Zartarian, G. "Piston Theory - A New Aerodynamic Tool for the Aeroelastician," *Jrnl. of the Aeronautical Sciences*, Vol. 23, No. 12, 1956, pp. 1109-1118.
- <sup>13</sup>Eckert, E.R.G., "Engineering Relations for Heat Transfer and Friction in High-Velocity Laminar and Turbulent Boundary-Layer Flow over Surfaces with Constant Pressure and Temperature," *Transactions of the ASME*, Vol. 78, No. 6, 1956, pp. 1273-1283.
- <sup>14</sup>Kennedy, M., O'Hagan, A., "Bayesian calibration of computer models". *Jrnl. of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 63, No. 3, pp. 425-464.
- <sup>15</sup>Kullback, S., Leibler, R.A., "On Information and Sufficiency", *Annals of Math. Stat.*, Vol. 22, No. 1, 1951, pp. 79-86.
- <sup>16</sup>Ryan, K.J., "Estimating expected information gains for experimental designs with application to the random fatigue-limit model," *Journal of Computational and Graphical Statistics*, Vol. 12, 2003, pp. 585-603.
- <sup>17</sup>Liang, B. and Mahadevan, S., "Error and Uncertainty Quantification and Sensitivity Analysis in Mechanics Computational Models," *International Journal for Uncertainty Quantification*, Vol. 1, No. 2, 2011, pp. 147-161.
- <sup>18</sup>Crowell, A. R., McNamara, J., & Miller, B., "Hypersonic Aerothermoelastic Response Prediction of Skin Panels Using Computational Fluid Dynamic Surrogates," *Journal of Aeroelasticity and Structural Dynamics*, Vol. 2, No. 2, 2011.
- <sup>19</sup>Neal, R.M., "Slice Sampling," *Annals of Statistics*, 2003, pp.705-741
- <sup>20</sup>Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., & McFarland, J. M., Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal*, Vol. 46, No. 10, 2008, pp. 2459-2468.
- <sup>21</sup>Bichon, B. J., McFarland, J. M., & Mahadevan, S., Efficient surrogate models for reliability analysis of systems with multiple failure modes. *Reliability Engineering & System Safety*, Vol. 96, No. 10, 2011, pp. 1386-1395.
- <sup>22</sup>Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., & Kim, N. H., Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, Vol. 132, No. 7, 2010.
- <sup>23</sup>Bect, J., Ginsbourger, D., Li, L., Picheny, V., & Vazquez, E., Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, Vol. 22, No. 3, 2012, pp. 773-793.
- <sup>24</sup>Basudhar, A., & Missoum, S. (2010). An improved adaptive sampling scheme for the construction of explicit boundaries. *Structural and Multidisciplinary Optimization*, 42(4), 517-529.
- <sup>25</sup>Chaudhuri, A., Le Riche, R., & Meunier M., Estimating Feasibility Using Multiple Surrogates and ROC Curves, *9th AIAA Multidisciplinary Design Optimization Specialist Conference*, Boston, MA, April 8-11, 2013
- <sup>26</sup>Belis, M. & Guillas, S., A quantitative-qualitative measure of information in cybernetic systems (Corresp.). *IEEE Transactions on Information Theory*, Vol. 14, No. 4, 1968, pp. 593-594.
- <sup>27</sup>Luan, H., Qi, F., Xue, Z., Chen, L., & Shen, D., Multimodality image registration by maximization of quantitative-qualitative measure of mutual information. *Pattern Recognition*, Vol. 41, No.1 , 2008, pp. 285-298.
- <sup>28</sup>Kennedy, M., O'Hagan, A., Bayesian calibration of computer models. *Jrnl. of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 63, No. 3, pp. 425-464.
- <sup>29</sup>Anderson, J. D., Jr., *Hypersonic and High-Temperature Gas Dynamics*, 2nd ed., AIAA, Reston, VA, 2006
- <sup>30</sup>Culler, A.J. and McNamara, J.J. (2010). "Studies on fluid-thermal-structural coupling for aerothermoelasticity in hypersonic flow." *AIAA J.*, 48(8), 1721-1738.
- <sup>31</sup>Rasmussen, C.E., Williams, C.K.I., *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.
- <sup>32</sup>Kullback, S., Leibler, R.A., On Information and Sufficiency, *Annals of Math. Stat.*, Vol. 22, No. 1, 1951, pp. 79-86.
- <sup>33</sup>Jones, D. R., Perttunen, C. D., & Stuckman, B. E., Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, Vol. 79, No. 1, 1993, pp. 157-181.
- <sup>34</sup>McNamara, J. J., & Friedmann, P. P., Flutter boundary identification for time-domain computational aeroelasticity. *AIAA Journal*, Vol. 45 No. 7, 2007, pp. 1546-1555.
- <sup>35</sup>Perez, R. A., Smarslok, B. P., & McNamara, J. J., Investigating Model Uncertainty in the Nonlinear Aeroelastic Response of Thin Panels. *17th AIAA Non- Deterministic Approaches Conference*, Kissimmee, FL, 2015.
- <sup>36</sup>Ferson, S., Oberkampf, W.L., & Ginzburg, L., Model Validation and Predictive Capability for the Thermal Challenge Problem. *Computer Methods in Applied Mechanics and Engineering*, Vol. 197, No. 29, 2008, pp. 2408-2430.
- <sup>37</sup>Jordan, A., On Discriminative vs Generative Classifiers: A Comparison of Logistic Regression and Naïve Bayes. *Advances in Neural Information Processing Systems*, 14, 841, 2002
- <sup>38</sup>Mitchell, T.M., *Machine Learning*, McGraw Hill, 1997